Evaluation Protocol

Thinking Maths – A professional learning program supporting middle-school teachers to engage students in Maths

23 September 2016

Evaluators

Australian Council for Educational Research
Prepared by Dr Hilary Hollingsworth and Dr Katherine Dix from the Australian Council for Educational Research.

Evaluation Summary	
Year levels	Years 6-9 (middle school)
Number of students	6500+
Number of schools	120 (+30 if recruitment drive attracts more than 120)
Design	RCT design – involves 60 intervention and 60 (+30) control schools, with sufficient power to achieve a detectible minimum effect size of 0.2
Primary outcome	The Thinking Maths intervention is designed to build teacher capacity to make mathematics learning deeper and more engaging, and in turn, improve student mathematics achievement outcomes. The evaluation will estimate the impact on achievement and cost per student for the program.

Contents

Background

Intervention name

Thinking Maths: Supporting middle-school teachers to engage all students more deeply in learning maths.

Rationale and essential elements

Thinking Maths is designed to build teacher capacity to make mathematics learning deeper and more engaging, and in turn, improve student mathematics achievement outcomes. The program is an initiative of the South Australian Department for Education and Child Development (DECD). Thinking Maths is a professional learning program to develop the pedagogical content knowledge of teachers for Years 6-9 and to foster consistent practices at the primary-secondary transition stage. The program aims to promote:

Primary Outcome:
- Improved student achievement in mathematics for all learners

Secondary Outcomes:
- The development of students as powerful learners of mathematics
- A shift in teachers' pedagogy towards more inclusive, student-centred learning (DECD, 2016a, p.2)

The key performance indicators identified by DECD for teachers as a result of participating in the program, will be:
- better mathematics content knowledge and more inclusive student-centred pedagogical knowledge and a repertoire of practices,
- clarity of beliefs about effective mathematics teaching and learning, and
- strengthened professional identity as teachers of mathematics.

The key performance indicators identified by DECD for Year 6-9 students in those classes, will be:
- greater development as powerful learners of mathematics (self-efficacy, cognitive engagement in mathematics learning and meta-cognition), and
- improvement in achievement above normal learning growth.

Materials and resources required for delivery

The materials and resources required for the delivery of the professional training days include resources that are used in the workshop sessions (training materials) and resources that participants are given to use back in their classrooms (teacher resources).

Training materials

- PowerPoint presentation Professional Learning (PL) slides (1 set per session)
- Large pencil case and materials per table - pens, scissors, tape, glue, ruler, highlighters, string, grid paper, sticky notes, etc. (6 reused)
- Various maths activities and equipment - blocks, chips, pan-balances, etc.
- Whiteboard, coloured markers
- Catering (morning tea, lunch for 5 days)

Participant (Teacher) resources

- A resource folder per participant (30) (online –TBA)
- Main textbook: Van de Walle (2014) Teaching Students-Centred Mathematics
- Handouts (e.g. believe scale sheet)
- Online resources and media (e.g. Jing, Jo Boaler – Brain Plasticity, National Library of Virtual Manipulatives, Solve Me Mobiles)
- AC Leaders Resource
- Edmodo login to Thinking Maths Group

What: Procedures, activities and processes

Year 6-9 teachers participate in professional learning and receive resources and ongoing support throughout the program, with the commitment that between sessions they reflect on and apply program ideas in their mathematics classes. It is anticipated that changes in teachers' practices, beliefs and self-efficacy will in-turn influence students' mathematics self-efficacy, cognitive engagement in learning, and meta-cognition, with the outcome of improved learning. Figure 1 presents the logic model of the Thinking Maths program, and specific details of the program follow.
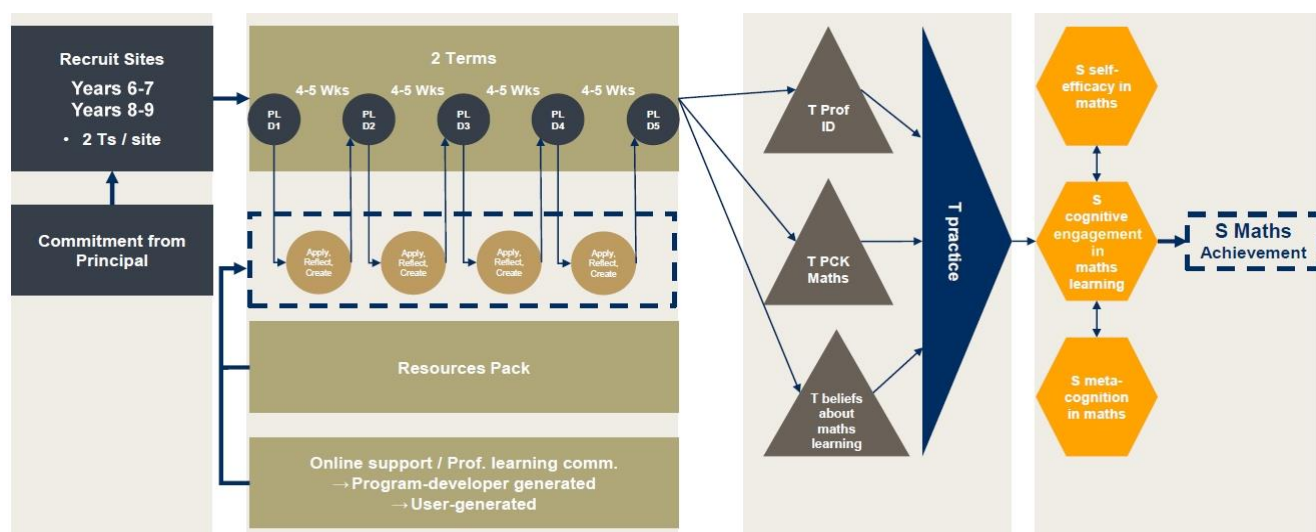

Figure 1. Conceptual model of the Thinking Maths program

Selection

Principals (of the targeted partnerships for our pilot program) will be contacted by DECD via an email inviting them to nominate at least two eligible teachers to be part of the program and to complete and send back an Expression of Interest form. Decisions about how this information is disseminated to teachers and mathematics leaders will be made at each site. Often the sites send a joint Expression of Interest for both teachers relating to site/ partnership priorities which may or may not have been written by the teacher(s). In previous programs there have been examples of the Principal making the decision for teachers to participate, and other examples where teachers hear about the opportunity through their networks and initiate the application.

PL Activities

Each of the PL sessions has a particular content focus and a particular pedagogical practice focus (see Table 1), and involves authentic classroom experiences with consideration for the perspective of individual learners.

Table 1. Professional learning content and pedagogical foci

PL Day	Content Focus	Pedagogical Focus
1	Patterns and Generalisation	Differentiating Learning
2	Space and Measurement	Effective Questioning
3	Geometry	Active and Collaborative Learning
4	Statistics	Personalising and Connecting Learning
5	Integers and Fractions	Teaching for Understanding

Each PL session runs for one full day and is structured to include the following collaborative activities, as outlined by DECD (2016a, pp.1-2):

- *Reflecting:* A reflection of a prior reading related to the day's foci. This may be from a professional journal or from the program's reference book, *Teaching Student-Centred Mathematics* (Van de Walle, 2014).

- *Sharing:* A sharing time when participants discuss the pedagogical strategy they trialled in their classrooms, including student responses and impact on student learning and dispositions. During this time participants show student work samples and present new lesson ideas.

- *Modelling:* Participants undertake mathematical activities presented by facilitators who model effective practices.

- *Learner/practitioner:* Participants engage in reflection about the workshop activities in regard to (i) their own learning experiences, content and processes, and (ii) the perspective of two of their mathematics students (i.e. as reflective practitioners).

- *Applying:* Participants map the workshop activities to curriculum documents, content descriptions, proficiencies and general capabilities. They consider the learning continuum to ensure all students have access to the curriculum and experience challenge and academic growth.

- *Accessing resources*: The identification and demonstration of video clips, websites, technology, provocations and other resources that can be used to engage and challenge students and/or use with colleagues. The repository of resources are designed to:

  o focus on building teacher conceptual knowledge of the content of the Australian Curriculum: Mathematics;

  o strengthen the repertoire of ideas and activities for mathematics teaching and learning consistent with a learning design approach to implementation of the *South Australian Teaching for Effective Learning Framework*;

  o include exemplars of student work (Every participating school will document evidence of the learning progressions for at least four students that can be shared. This could include annotated work samples; assessment pieces showing evidence of moderation with other colleagues; problem-solving tasks showing the application of student-led learning design; learning plans and programs showing how students' learning experiences are modified based on feedback and responses by students; etc); and

  o include an annotated facilitator's folder to support numeracy leaders to provide professional learning in their own context.

Processes adopted by teachers

The "high-gain strategies" listed below, are explicitly demonstrated during the PL sessions. Participants are asked to implement these back at their schools and within their classroom (DECD, 2016b, pp.2-3). This process is strengthened when two teachers from a school can work together, hence the selection criteria for participation in the professional learning program.

- Participate in a professional learning community in your school and partnership.
  - 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 students 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.

The program asks teachers to identify and work on aspects of the high-gain strategies that are most appropriate and achievable for them, their students and their context (DECD, 2016b, pp.2-3). Teachers are asked to 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.

Activities that teachers are expected to undertake outside of the professional learning days include:

- undertake required professional reading
- utilise the reference book (Van de Walle, 2014) where appropriate in their learning design
- participate in the online discussions
- trial strategies and tasks in their own classrooms
- share resources and as appropriate upload these to the PLC (professional learning community) forum
- collect and analyse student work samples
- keep a professional journal to support the sharing of ideas at the PL days
- share their learning with co-participant colleagues at their site and in their partnership

Who: Expertise and background

The five days of PL and two terms of online support are conducted by two DECD professional facilitators, Dr Pauline Carter and Ms Maureen Hegarty. Both are highly experienced middle-school mathematics teachers who have extensive experience in teacher professional development and pre-service teacher training.

Dr Pauline Carter holds a PhD in Education and is Project Officer, Critical and Creative Thinking Strategic Division, Learning Improvement Division at DECD. Ms Maureen Hegarty holds a Masters of Education and is a teacher and lecturer in Maths Education. They are the co-developers of the Thinking Maths program and also the co-facilitators. Short biographies for these facilitators are presented in the Personnel section.

How: Modes of delivery

Two delivery modes are utilised in the Thinking Maths program: face-to-face delivery and online support.

Five days of face-to-face delivery are provided to groups of 30 participants at a time, by the two qualified facilitators. Topics are presented to the whole group, along with sharing and discussion. Participants also work collaboratively on maths activities in small groups of five or six around tables.

An online professional learning community, using the Edmodo platform, is moderated by the professional facilitators. Participants from all groups are invited to contribute to this community.

Where: Location and infrastructure

Thinking Maths is a program designed to be inclusive of all upper-primary and lower-secondary mathematics teachers in South Australian government schools. Accordingly, teachers can come from metropolitan, regional and remote locations. During this pilot phase, the PL training sessions are centrally located and held at the Education Development Centre in Hindmarsh, South Australia. The PL events require a conference/meeting room big enough for seven tables (30 people), presentation facilities with speakers and internet, and easy parking and access.

When and how much: Dose

At the heart of Thinking Maths is five days of face-to-face professional learning that is programmed over two terms at 3-4 week intervals.

- Learning: Five PL days, 3-4 weeks apart, generally over two terms – Day1, Wk8 Term1; Day2, Wk2 Term2; Day3, Wk8 Term2; Day4, Wk2 Term3; Day5, Wk6 Term3. In total, the program involves 30 hours of face-to-face professional learning, with an additional expectation of engagement in reading, journaling and presenting to the group.

- Implementation: Teachers use the four periods of 3-4 weeks in between the PL days to reflect on and apply program ideas in their mathematics classes. This implementation process follows a cycle of Action, Reflection and Creation. Teachers' journal and share the evidence of changing behaviours and outcomes in subsequent PL sessions.

- Support: Ongoing support and participation in online professional learning community (Edmodo)

An indicative breakdown of the time involved by teachers and exposure to students is presented in Table 2.

Table 2. Teacher and student dosage

Participant	Activities	Time
Teacher	PL sessions (5 x 6hrs)	30 hrs
	Lesson preparation (per week)	2 hrs/wk
	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)	60 hrs/wk
	Secondary students in maths class (per week)	30-48 hrs/wk

Significance of the evaluation

Thinking Maths has been designed to address the drop in maths performance between Years 7 and 9 identified in NAPLAN results. It occurs at a time of significant change in students' lives as they transition from primary to secondary school and undergo social, physical and emotional changes (Redmond et al., 2016). The focus of the program is on the teachers of students in these transition years.

The aim of this efficacy evaluation is to find out if the Thinking Maths program works to improve the mathematics attainment of South Australian students, particularly those from disadvantaged backgrounds. To achieve this, the Thinking Maths intervention will be subject to a quantitative impact evaluation to estimate the effect on student's maths attainment, primarily. An implementation and process evaluation is also being undertaken to understand how the project is implemented on the ground and the elements of successful delivery. This is being carried out by ACER, as an independent evaluator to ensure that the evidence produced is robust and unbiased.

The program uses professional development strategies that have a strong evidence base. The pedagogical principles of the program are research-based and draw on the work of leading educational experts such as Sullivan (2011, 2013), Dweck (2000), Claxton (2012), Boaler (2015, 2005, 2002) 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 Australian Curriculum Mathematics in differentiated ways responsive to individual student's needs, interests and dispositions.

For example, regularly sharing teaching experiences and discussing what works and doesn't work with colleagues, supports improved teaching practice, as does the opportunity for teachers to become the learner to increase the visibility of learning from the students' perspective (Miller, 2009). Another example of a strategy used, is to first address students' low self-efficacy before trying to raise their achievement (Miller, 2009). Teachers are also encouraged to adopt a Growth Mindset as described by Dweck and colleagues (Dweck, 2000; Dweck et al., 2014). Their work highlights links between confidence and self-efficacy in mathematics, mastery of problems, and building resiliency, when teachers and students work together. A further strategy involves the PL 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).

Although this project was first trialled in 2014 with two classes, and participating teachers were surveyed to provide their feedback to the professional learning experience (including consequent shifts in pedagogy and observed changes in student engagement and achievement), student achievement data was not collected. Quantitative student outcomes were considered or available (i.e., NAPLAN was not sufficiently timely and PATMaths test were not introduced state-wide until 2015).

Specifically, the evaluation in 2017 will constitute an efficacy trial of a five-day professional learning program for Years 6 to 9 mathematics teachers, in order assess the beneficial effects of the program under optimal conditions of delivery (Flay et al., 2005). The program supports two teachers from each site in the deep learning of mathematical content as outlined in the Australian Curriculum Mathematics. In particular, it focuses on quality task design, the sequencing of conceptual development and research-informed effective pedagogies for the teaching and learning of mathematics. The evaluation will involve 120 teachers from 60 schools, divided into four concurrent groups of 30 teachers in each group. This assumes that the developers of the Thinking Maths program (Pauline and Maureen) will co-present and deliver concurrently the four lots of five professional learning sessions – 20 sessions in all. The efficacy trial will evaluate student achievement by using PATMaths student data routinely collected by DECD in 2016 and 2017.

Methods

This funded project is an evaluation that will be conducted in 120 South Australian government primary and secondary schools using a Randomised Controlled Trial (RCT) research design with pre- and post-intervention assessments. Our principal aims will be to measure the extent to which Thinking Maths processes enable middle-school teachers to implement the Thinking Maths goals and strategies (why and how it works), and the impact that these processes, goals and strategies have on student mathematics outcomes (if it works).

Research questions

In order to ensure that the measurement framework (see Analysis Plan section) and resulting data collection tools are effective, it will be important to link the evaluation questions to outcomes from the outset (Radhakrishna & Relando, 2009). Figure 2 provides a roadmap of the key components involved in linking evaluation questions to program outcomes. ACER has worked closely with DECD in this important preparation stage to ensure that the study is underpinned and guided by appropriate evaluation questions.
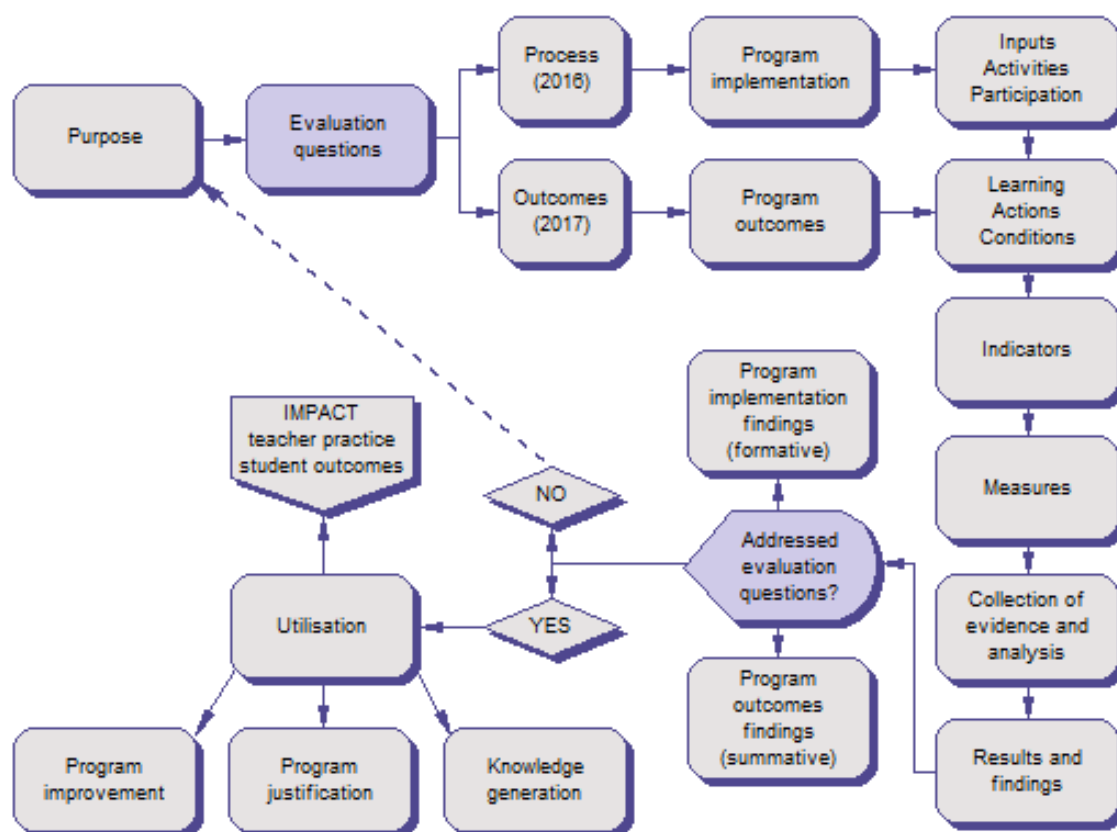

Figure 2. Schematic description of linking evaluation questions to program outcomes and impacts

The broad research question that will be addressed is: To what extent does the SA DECD Thinking Maths program improve student mathematics outcomes, and build teacher capacity to make mathematics learning deeper and more engaging?

Outcomes Questions – Is the program accomplishing its intended results?

Based on the evaluation design, we will use the Thinking Maths Implementation Index (discussed below) to address twelve research questions. The following are five research questions specific to outcomes.

1. Did the Thinking Maths program enable middle-school students (Years 6-9) to improve their mathematics achievement (PATMaths - Progressive Achievement Tests in Mathematics scores) above typical learning growth?

2. How did Thinking Maths develop middle-school students (Years 6-9) as powerful learners of mathematics in terms of:
   a. mathematics self-efficacy,
   b. cognitive engagement in learning, and
   c. meta-cognition?

3. How 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 (e.g., self-efficacy)?

4. How did Thinking Maths shift teachers' mathematics teaching practice towards a more inclusive, student-centred learning approach?

5. Did changes in teachers' practices due to Thinking Maths, influence student's mathematics outcomes?

Implementation and process questions - How is the program operating?

We recognise that an understanding of the implementation and process evaluation is essential at all stages of intervention. The process data will 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 Thinking Maths works). This approach is underpinned by the multi-disciplinary field of implementation science, which broadly aims to enhance the effectiveness of interventions by understanding them. The field recognises that there are multiple dimensions that affect implementation (Humphery et al., 2016), but commonly involve:

- Fidelity/adherence–the extent to which implementers (e.g. teachers) adhere to the intended treatment model
- Dosage–how much of the intended intervention has been delivered and/or received
- Quality–how well different components of an intervention are delivered
- Reach –the rate and scope of participation
- Responsiveness – the degree to which participants engage with the intervention

- Programme differentiation–the extent to which intervention activities can be distinguished from other, existing practice
- Monitoring of control/comparison groups (in a trial context) –determination of the 'counter-factual' (e.g. that which is taking place in the absence of the intervention)
- Adaptation–the nature and extent of changes made to the intervention

As a basis for understanding changes in mathematics outcomes, we will create a Thinking Maths Implementation Index using the quality, fidelity and dosage of the implementation of the program, in order to examine the effects of the program on teachers and students. However, we also recognise that there are three levels involved in the implementation of Thinking Maths, which will need to be considered. The first level involves the PL facilitators and the content itself, the second level involves teachers and their level of engagement, internalisation and reflection, and the third level involves the students and what happens back in the classroom. Understanding what is happening at each level will be important when it comes to explaining why or why not there is an impact. This will be particularly important, if there is no impact, in explaining at what level a failure of the process may have occurred.

With this in mind, we propose the following seven process research questions.

6. What are the critical elements of the Thinking Maths program, in terms of quality of delivery, fidelity and dosage?
7. How applicable and useful is the Thinking Maths approach (PL, online community, support, resources) in primary and secondary school settings?
8. To what extent did teachers engage with the Thinking Maths program?
9. How cost-effective is the Thinking Maths program?
10. What are the barriers and facilitators to the effective implementation of Thinking Maths in middle-school classrooms in different contexts (Year level, school socio-economic status, location, high proportions of Indigenous students)?
11. How can the Thinking Maths program be improved?
12. What are the risks and challenges in expanding the Thinking Maths program to scale?

Design

This proposed evaluation uses a RCT block randomisation design, with quantitative pre- and post-data collection including PATMaths tests alongside online surveys with teachers and students. This data, collected by ACER, will be augmented with existing sources of data provided by DECD and participating schools in the form of administrative/enrolment records after appropriate permissions have been obtained. The overarching approach aligns with recommendations made by Torgerson and Torgerson (2013), adapted in Figure 3.
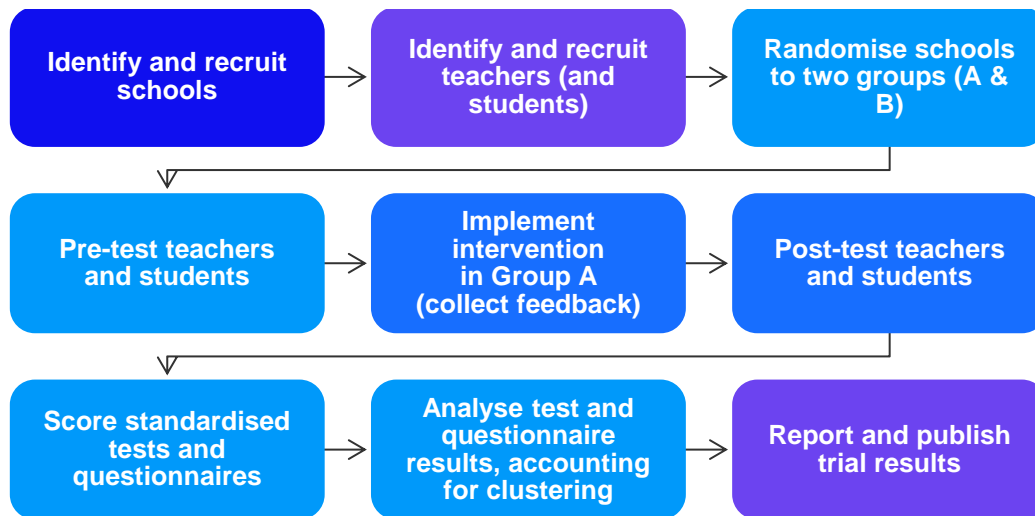
Figure 3. Key steps for a successful school-based RCT

The proposed design is based on a two armed (intervention and control) clustered RCT with block stratified randomisation at the school level, as shown in Figure 4. At least 120 schools will be recruited through a self-selection process by submitting an Expression of Interest. These schools will be randomly assigned to the intervention (Group A) and the control (Group B). Group A schools will commence in Term 1 2017, while Group B schools will first act as control and have a delayed start, commencing in Term 4 2017.
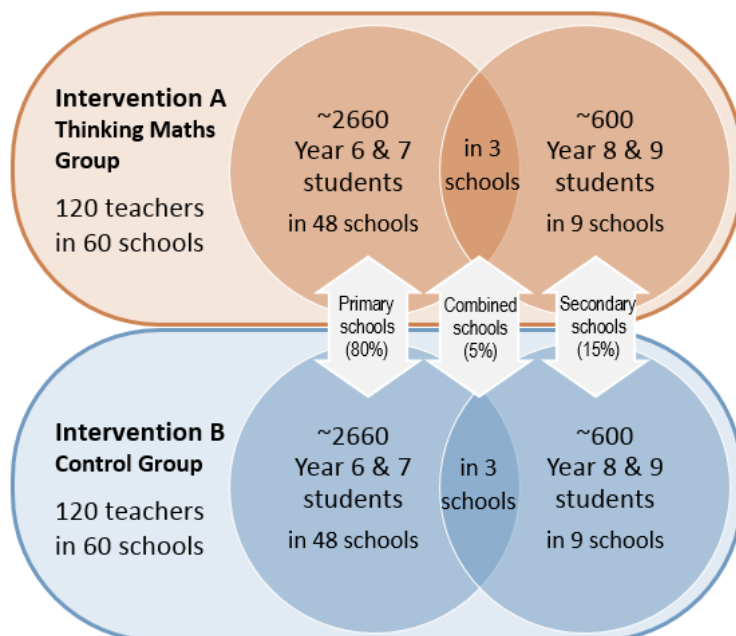

Figure 4. RCT evaluation design: schools, teachers, students

Randomisation

This study will used concealed randomisation so that there is no foreknowledge of the randomised allocation (Torgerson & Torgerson, 2013). Randomisation of schools will be done after teachers have been recruited and given their consent to participate in the study. Accordingly, all participants, including teachers, schools and DECD (the recruiters and program implementers) will not know which group the schools are randomised into until after this has happened.

In order to maintain independence and concealment from the program implementers (DECD) and the evaluation funders (SVA), DECD will provide the sampling frame of participating schools to ACER, and ACER will undertake the randomisation process using robust methods involving the SPSS computer program. A stratified approach using school type will achieve balanced representation of primary and secondary schools in the implementation and control groups, as Figure 4 suggests. The randomisation process used by ACER will be document in Progress Report 2.

Participant eligibility

Recruitment through an Expression of Interest is undertaken at the school level by DECD for 150 sites, where 60 sites receive the intervention (Group A) and 90 sites act as control (Group B). Group B includes up to 30 additional school to allow for control attrition. It is preferable to have two teachers from each school receiving the intervention (120 teachers), but not to the exclusion of small schools.

Eligible schools need to meet the following criteria:

- Government school located in South Australia
- Cater for students in Years 6-7 and/or Years 8-9 (K-12 Area schools are counted as one site)
- School have not previously received the Thinking Maths (or equivalent) intervention
- Teachers teach a Year 6, 7, 8 and/or 9 class in mathematics and have not previously participated in the Thinking Maths (or equivalent) intervention

Primary outcome measure – Maths Achievement

The primary outcome identified in this evaluation – the outcome that determines whether or not the intervention is effective – is improved student achievement in mathematics for all learners. This will be measured by the standardised PATMaths test routinely completed by all students in South Australian government schools since 2015.

The ACER Progressive Achievement Tests in Mathematics (PATMaths) Fourth Edition (2013) 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. Each of the ten PATMaths tests assesses the content of

one year level of the Australian mathematics curriculum from Year 1 to Year 10, which assumes coverage of the curriculum of lower year levels.

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 to be completed by students online within 40 minutes.

In accordance with DECD's mandatory annual administration of PATMaths, the pre-test data will be collected during September 2016, with data retrieved retrospectively once the schools, teachers and their students participating in the Thinking Maths evaluation have been identified in early 2017. The post-test data collection will occur in September 2017.

The test is scored instantaneously through the ACER Test Scoring and Analysis software. Schools are automatically provided with their students' raw score, Rasch scaled score and infit, through a suite of group and individual reports. These interactive reports can be generated immediately after the tests are completed online and compared to previous years for the assessment of longitudinal growth.

ACER will also have access to these reports, enabling the retrieval and compilation of student-level pre- and post-data. The resulting database will be coded and de-identified once matching of the 2016 and 2017 data has been undertaken.

In addition to student name, Year-level, class group and school, the data will comprise the following results for each student.

- The test 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.

- The percentile rank for a particular test raw score shows the percentage of the students tested whose scores fell below that test raw score. It is a measure used to rank candidates in a reference sample, not a percentage score on the test.

For example, a test raw score of 18 on PATMaths Test Year 1 is equal to a scale score of 31.4, whereas the same test raw score on PATMaths Year 2 is equal to a scale score of 40.1. This example shows that relying on test raw scores alone does not give an accurate picture of a student's ability. Obtaining the same score on both tests could suggest that the two results are equivalent, whereas a comparison of the scale scores shows that the second score is much higher than the first (ACER, 2011). For the purposes of this evaluation, the scale scores will be used in the pre-post analysis.

Secondary outcome measures

The impact on student and teacher affective factors, such as engagement in mathematics and mathematics self-efficacy (as highlighted in the logic model, Figure 1), will be examined through questionnaire data gathered before and after the course of the intervention. Accordingly, the secondary outcomes identified in this evaluation and how they will be measured are:

- Students develop as powerful learners of mathematics, measured by a student attitudinal questionnaire.

- Shift in teachers' pedagogy for more inclusive, student-centred learning, measured by a teacher attitudinal questionnaire and artefacts collected as part of the professional learning sessions (e.g. shared stories, professional journal, online community).

In consultation with DECD, ACER will develop pre- and post-intervention teacher and student questionnaires that will be delivered online and take no more than 30 minutes to complete (within a lesson time). These will focus on the affective aspects of the program and be completed by intervention and control participants by selecting Likert-type response categories (e.g. strongly disagree – 1, to strongly agree – 5). Where possible, items will be sourced from existing psychometrically tested and validated scales (for example PISA 2012). Responses will be automatically scored and collated into a secure downloadable database through the online survey hosting platform.

In addition, ACER will develop PL feedback forms completed by intervention participants. In order to maximise response, forms will be limited to a single two-sided page and administered at the end of each PL session. DECD will collect the forms and provide them to ACER for digital data extraction using OCR form scanning software. Quality control measures will be used to minimise error in any hand-keyed data.

A range of additional background variables including student characteristics, teacher and classroom characteristics, and school level characteristics will be included on the questionnaires and forms or provided by DECD, for background information and identification purposes.

A final source of qualitative data that will inform the secondary outcomes and process evaluation are the discussions and sharing by participants that are documented during PL sessions, along with artefacts collected (e.g. shared stories, professional journal, online community discussions). The types of data that will be collected are summarised in Table 3 in the Analysis Plan section.

Sample size calculations

In order to detect an effect that is sufficiently large to be of educational significance at the student level (i.e. above 0.2), and given that teachers are clustered within schools, the following recommendations about sample size are provided.

We take the desired alpha to be 0.05 and power to be 0.8, with a detectable effect size of small (Cohen's d = 0.2). Even using the simplest formula for a RCT block design comparing two groups of equal size, we also need to take into account the design effect of clustering and intra-cluster correlations (ICC). Since students in one school are more like each other than students in another school (Hutchison & Styles, 2010; Torgerson & Torgerson, 2013), the sample is not a simple random sample, and results 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. The 'design effect' is used to estimate the extent to which the sample size should be inflated to accommodate for the homogeneity in the clustered data. In line with similar studies in Australia, we are adopting an ICC coefficient of ρ = 0.3, but will review this once data are available.

In order to minimise sample size and achieve the desired Minimum Detectable Effect Size (MDES) of 0.2, the MDES formula with both level-1 and level-2 covariates given by Bloom et al. (2007) is used, 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 + 30 additional if available) is needed to achieve a MDES of 0.2 with covariates that accommodate design effects and provide allowances for participant attrition and missing data.

Analysis plan

This section discusses in more detail the plan for how the analysis of the data collected will be defined and treated in order to address the research questions and the aims of the evaluation. It begins with general discussions about our approach to analysis and is followed by the measurement framework that summarises the specific aspects being measured and their treatment.

Data cleaning

The data will be thoroughly checked and cleaned to address any coding issues and to prepare the data in a systematic manner for analysis. Dubious data can influence adversely subsequent analyses, hence the data cleaning process aims to eliminate impossible or incorrect values. While the computer based design of the data capture should minimise these errors, data cleaning is still required to ensure data accuracy and consistency. Further, results for items that had validation rules will need to be checked to ensure the applicability and appropriateness of these rules.

Administration issues

Analysis of the pre-questionnaires will provide insights into the survey administration process. Careful examination of frequencies of each item and comparisons with the presentation of the questions as they appeared to respondents will provide insight into administration issues. While the questionnaires will be piloted to ensure that length is appropriate and completion time is reasonable, missing data at the end of pre-questionnaires may indicate that the post-questionnaire needs to be shortened.

Missing data

As with any data collection process, there will be missing data that may arise for several reasons, such as:

- The participant might chose not to answer an item or inadvertently missed an item.

- The participant might feel that a section of items is not relevant to them personally.

- Items from a paper-based feedback form may be missing due to data entry error.

- The questionnaire or feedback form is returned only partially completed, not completed at all, or not administered.

Missing data will be coded with a single missing code value (999) to represent all missing data. Methods will be used during scale score construction that avoid the need to impute missing data. Careful consideration will be given to the existence of missing data likely to bias the findings of the evaluation with respect to its representativeness.

Frequencies

For each item, frequencies of categorical questions and descriptive statistics for numerical questions will be calculated, and as they become available, will be reported in progress reports. These statistics - minimum, maximum, mean, range, mean and % missing - will be used to examine the behaviour and properties of each item in terms of their distribution, use of categories (e.g. all have been used) and proportion of missing data which should not exceed 10 per cent.

Scale analysis

The questionnaires and feedback forms will contain groups of items that will be designed to form a scale (e.g. Five items assessing mathematics self-efficacy). In order to consider the measurement of these scales, the classical item statistics for these scaled items will be reported. These will include the following:

- Scale reliabilities (Cronbach's alpha): Generally, reliabilities of 0.80 or more are described as high; between 0.70 and 0.80 as moderate; between 0.60 and 0.70 as low; and below 0.60 as very low.

- Item-total correlations: These correlations indicate to what extent individual items correlate with the overall scale score (for all other items). Low item-total correlations

- (< 0.3) indicate items with poor scaling properties.

- Numbers of respondents with valid and missing responses

- If the items are heavily skewed, the estimation of scale validity and reliability will require the use of distribution-free techniques in preference to using transformations to normalise the data. In order to assess the unidimensionality of scale constructs, confirmatory factor analysis will be carried out in AMOS (Arbuckle, 2007) using asymptotically distribution-free (ADF) estimation methods.

Factor analysis

While we will seek to use pre-existing scales, there may be need to develop new scales. Where necessary, exploratory and confirmatory factor analysis will be conducted in order to ensure the structural validity of the scale measures. Results of factor analyses may also be used to shorten scales, with option to revise the post-questionnaires by removing items that do not add much to the measurement of the underlying factor.

Correlation

Validity checks will be done mainly by way of correlation analyses to examine whether hypothesised relationships between scales hold. An example of a convergent validity check would be that higher mathematics self-efficacy would be related to higher mathematics

achievement, whereas an example of a discriminant validity check would be to see whether there is actually no relationship between student gender and teacher's professional identity.

Scale score construction

Following the assessment of the distribution characteristics of items, appropriate scaling methods will be employed depending on how normal or non-parametric the distributions are. It is not uncommon for attitudinal and perception data to be skewed and may require the use of non-parametric techniques such as confirmatory factor analysis using asymptotically distribution-free estimation, where models for each scale are built and tested and factor scores derived. If data are sufficiently normally distributed then standard factor-scores or mean response across the items will be derived using SPSS.

Implementation index

Given the complexity of implementing a program in schools, it is anticipated that some teachers will engage more readily than others with the Thinking Maths program, and so will be better able to effect change. In recognition of concerns about evaluating the quality of implementation (Durlak & DuPre, 2008), an Implementation Index will be developed. By doing so, it will strengthen the rigor of the evaluation and the ability to attribute the effects of the intervention to improved outcomes (e.g., Dix et al., 2012).

To identify schools (and teachers) as being 'high' or 'low' implementers of Thinking Maths, a framework derived from the work of Domitrovich et al. (2008) will guide the development of the Index that will represent the quality of schools/teachers implementation of Thinking Maths in terms of three elements: fidelity of implementation, extent of the dosage delivered, and the quality of the delivery process. The Index will be informed by data derived from the views of those experiencing the intervention (teachers and students), as well as those providing dedicated support for the implementation (Thinking Maths Facilitators).

The responses collected during the evaluation period will result in a character profile for the school and a score. Based on the profile and score, statistical modelling using Latent Class Analysis (Muthén & Muthén, 2007) will allocate the schools/teachers to low, medium or high groups. Once constructed, the Implementation Index can be used both as an outcome, to determine what factors influence implementation quality (e.g., SES), and also as a predictor, to determine under what conditions (level of quality) the Thinking Maths program has influenced outcomes (e.g., PATMaths gains).

HLM analysis

Because this evaluation involves whole classes, within schools, the students' outcomes are not independent of each other and any analysis will need to take clustering into account.

ACER will employ multiple approaches, such as treating each class as a single observation and comparing cluster-level means, through to using complex statistical modelling

approaches, such as hierarchical linear analysis, which takes into account the nested nature of the data (students within classes, within schools) while avoiding data reduction.

In order to test for significant relationships using a technique that takes into consideration the nested nature of the data and does not depend on assumptions of normality, two and three-level hierarchical linear models (HLM) will be built to assess each student and teacher outcome variable. Version 5 of the HLM program will likely be used in preference to more recent versions since it has greater capacity to handle missing data (Bryk and Raudenbush, 1992). Hierarchical linear analysis seeks to take into consideration the hierarchical nature of complex multilevel data, resulting from nested samples like the one used in this evaluation. In HLM, each level in the nested structure is formally represented by its own sub-model. Raudenbush and Bryk (1994, p.7) explain that "these sub-models express relationships among variables within a given level, and specify how variables at one level influence relations occurring at another". According to Raudenbush and Bryk (2002), the advantages that HLM had over single-level techniques, include its ability to improve the estimation of individual effects, to formulate and test for cross-level effects, and to partition variance and covariance components between levels of analysis. For these reasons, HLM will be used in this evaluation in order to give rise to models that are applied more meaningfully to the situation in which Thinking Maths is conducted.

As an example of the approach to analysis, Figure 5 presents a theoretical two-level hierarchical model of the factors influencing students' mathematics achievement, which also includes the Thinking Maths Implementation Index at the uppermost level. Full details about model specification will be documented in subsequent progress reports.
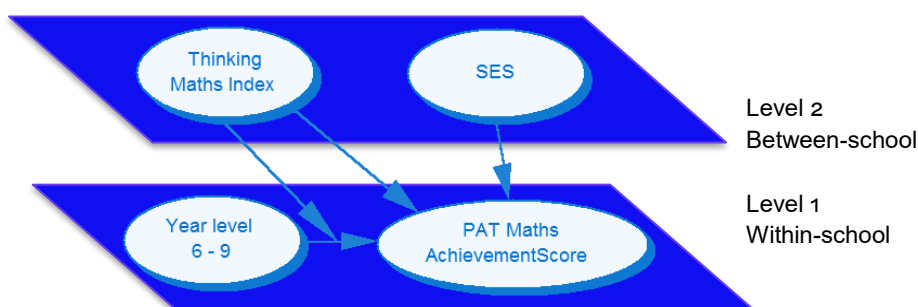
Figure 5. Theoretical two-level HLM model of factors influencing mathematics achievement

Qualitative data analysis

Discussions and sharing by participants documented during PL sessions, along with artefacts collected (e.g. shared stories, professional journal, online community discussions) will be analysed using three approaches based on:

a. Program usability and applicability involving an analysis of facilitators and barriers to implementation.

b.   Thematic analysis, examining the emerging themes from and applied to teaching practice.

c.   Exemplars and contextual influences, to identify examples of good implementation.

A summary of the emerging themes into core messages relating to Thinking Maths implementation will be provided.

Measurement framework and analysis plan

Taking into consideration the above general discussion, Table 3 presents the measurement framework and outlines the analysis plan for the Thinking Maths evaluation. It maps the data sources and the approaches to analyses against the conceptual factors that underpin the research questions. The plan uses a mixed methods approach of quantitative analyses and data modelling, triangulated with qualitative thematic analyses in order to provide a representative, but details insight into the Thinking Maths program. The items and scales presented in Table 3 are indicative and will be finalised in consultation with DECD and SVA during the questionnaire development stage.

Table 3. Thinking Maths evaluation measurement framework and analysis plan

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Primary Outcome					
Students' mathematics achievement scores	1	Online Student PATMaths	Student ID (name) Year level Maths achievement: number, algebra, geometry, measurement, statistics and probability	Rasch scaled score	Individual student scale scores Pre/post effect size
Secondary Outcomes					
Students' mathematics self-efficacy	2a, 5	Student questionnaire e.g. PEEC 2016	a) I usually do well in mathematics b) I am just not good at mathematics* c) I learn things quickly in mathematics d) I am good at working out difficult mathematics problems e) Mathematics is harder for me than any other subject* f) I would like to do more mathematics in school	Likert-type scales SD(1) – SA(4)	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Students' cognitive engagement in learning	2b, 5	Student questionnaire e.g. PEEC 2016	a) I know what my maths teacher expects me to do b) My maths teacher is easy to understand c) I am interested in what my maths teacher says d) My maths teacher gives me interesting things to do	Likert-type scales SD(1) – SA(4)	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM
Students' cognitive Engagement in learning	2b, 5	Student questionnaire e.g. PISA 2012 (note that these example items are for 15 year olds)	a) The teacher asks questions that make us reflect on the problem b) The teacher gives problems that require us to think for an extended time c) The teacher asks us to decide on our own procedures for solving complex problems d) The teacher presents problems for which there is no immediately obvious method of solution e) The teacher presents problems in different contexts so that students know whether they have understood the concepts f) The teacher helps us to learn from mistakes we have made g) The teacher asks us to explain how we have solved a problem h) The teacher presents problems that require students to apply what they have learned to new contexts i) The teacher gives problems that can be solved in several different ways	Likert-type scales Never Sometimes Often Always	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM
Students' meta-cognition	2c, 5	Student questionnaire e.g. MSLQ	a) When I become confused about something in maths, I go back and try to figure it out b) During math lesson I often miss important points because I'm thinking of other things	Likert-type scales	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Students' meta-cognition	2c, 5	Student questionnaire e.g. PISA 2012 (note that these example items are for 15 year olds)	a) The teacher sets clear goals for our learning b) The teacher asks me or my classmates to present our thinking or reasoning at some length f) The teacher asks questions to check whether we have understood what was taught h) At the beginning of a lesson, the teacher presents a short summary of the previous lesson l) The teacher tells us what we have to learn	Likert-type scales Never Some lessons Most lessons Every lesson	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM
Teachers' pedagogical and content knowledge	3a, 5	Teacher questionnaire	When teaching maths, how often do you a) Work together to try out new ideas b) Collaborate in planning and preparing instructional materials c) Discuss how to teach a particular topic d) Share what I have learned about my teaching experiences	Likert-type scales Never to Almost daily	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM
Teachers' beliefs about mathematics teaching & learning	3b, 5	Teacher questionnaire	Instructional practices a) use a variety of assessment strategies b) provide appropriate challenges for very capable students c) implement alternative strategies in your classroom d) gauge student comprehension of what you have taught e) adjust your lessons to the proper level for individual students f) provide an alternative explanation when students are confused g) respond to difficult questions from your students h) craft good questions for your students	Likert-type scales	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Teachers' professional identity (e.g., self-efficacy)	3c, 5	Teacher questionnaire	Student engagement a) get through to the most difficult students b) assist families in helping their children do well in school c) help your students think critically d) establish routines to keep activities running smoothly e) improve the understanding of a student who is failing f) motivate students who show low interest in schoolwork g) foster student creativity help your students value learning h) get students to believe they can do well in schoolwork	Likert-type scales None at all Very little To some degree Quite a bit A great deal	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM
Teachers' mathematics teaching practice (more inclusive, student-centred learning approach)	4, 5	Teacher questionnaire	When teaching maths, how often do you a) Bring interesting materials to class b) Relate the lesson to students' daily lives c) Summarise what students should have learned from the lesson d) Encourage all students to improve their performance e) Use questioning to elicit reasons and explanations f) Praise students for good effort g) Let students direct their learning	Total scale score Never to Almost daily	Descriptive stats, distributions - SPSS Cronbach α Item reliability analysis HLM Thematic analysis of artefacts
Process					
Program quality	6	Observation PL Feedback Teacher questionnaire	Fidelity - Degree to which an intervention is conducted as planned Dosage - Specific units of an intervention and support system Quality of delivery - Affective engagement with the process and support responsiveness	Likert-type scales	Implementation Index – school-level score indicating the 'quality of engagement' based on

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Program applicability and usefulness	7	Teacher questionnaire PL Feedback	At each level Level 3: Facilitators - delivery Level 2: Teachers - reflection Level 1: Students - outcomes For each component (PL sessions, online community, support, resources)	Likert-type scales	identified indicators. Latent class analysis (MPlus) to assign schools into low, medium, high engagement
Teachers' engagement with program	8	Teacher questionnaire PL Feedback	Attend all session Actively participated in online discussion Completed readings Reflective journal Sharing evidence of class work	Scores	
Cost data	9	DECD	Training materials - Powerpoint presentation PL slides (1 set/session) Large pencil case and materials per table - pens, scissors, tape, glue, ruler, highlighters, string, grid paper, sticky notes, etc (6 reused) Various maths activities and equipment - blocks, chips, pan-balances, etc Whiteboard, coloured markers Catering. Venue hire. Participant (Teacher) resources, resource folder, text Van de Walle (2014), Handouts (eg believe)	Set-up costs Running costs	Cost per student
Facilitators and Barriers	10	Teacher questionnaire PL Facilitators	a) What have been the barriers to implementing Thinking Maths with your class? What is an example of this in practice? b) What has facilitated the implementation of Thinking Maths with your class? What is an example of this in practice?	Open text	Thematic analysis Sub-group analysis
Improvement	11	PL Facilitator PL Feedback	How can the Thinking Maths program / this session be improved?	Open text	Thematic analysis
Scaling up	12	PL Facilitator Interviews	What are the risks and challenges in expanding the Thinking Maths program to scale?	Open text	Thematic analysis

Conceptual domain	RQs	Data source	Items or examples	Codebook / Output	Analysis approach
Context					
School background		DECD / School data / MySchool site	Type (P/S/Comb) Loc (metro/reg/remote) SES (ICSEA) ATSI%	Numeric	Descriptive stats, distrib - SPSS
Teacher background		Teacher questionnaire	Gender Teaching experience Teaching in specialisation	Numeric	Descriptive stats, distrib - SPSS
Student background		Student questionnaire / DECD	Gender Year-level / Age ATSI status	Numeric	Descriptive stats, distrib - SPSS

Cost data

It is increasingly recognised that collecting cost data alongside and contemporaneously with the effectiveness evidence is important for decision-makers (Belfield et al., 2013; Levin et al, 2012). In education, a cost-effectiveness analysis supports evidence-based decision-making to ensure that the limited resources and funds given to a program are used in the most efficient and cost-effective way, while maintaining program objectives (Levin & Garcia, 2013; Hummel-Rossi & Ashdown, 2002).

Our cost analysis follows the 'ingredients method' (Chambers & Parrish, 1994; Levin, 1995), which accounts for the costs of the resources required to implement the educational intervention being evaluated, rather than focusing on a budget. The approach involves "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 will be costed, including TRT time, facilitator costs, materials, venue hire, administration, etc.

ACER, with assistance from DECD, will establish a systematic specification of the 'ingredients' that are used in setting up and running the Thinking Maths program, to which prices will be assigned, in order to determine the cost of the intervention. This information will be compiled by DECD using a proforma. The basic technique (Rossi & Freeman, 1985) combines this cost data with the results for educational effectiveness, in our case, PATMaths gain scores. Accordingly, the cost-effectiveness ratio is based upon the average effects and costs per student. By dividing the number of students of the teachers participating in the intervention by the overall cost, an estimate of cost per student can be derived and the cost-effectiveness ratio established. A similar analysis will be undertaken with the control schools (Group B) in order to establish a 'business-as-usual' cost comparison. This information will also be supplied by DECD near the end of the evaluation period.

Ethics and registration

Ethical approval for the research undertaken in this study with teachers and their students in Years 6 to 9, must first be gained by the ACER Ethics Committee. An ethics request was submitted and approval was gained on 30 June 2016 (Ref no. 544883).

Project approval must also be gained from DECD to undertake research in South Australian government schools. This may involve completing their application, consent form and checklist to conduct research. Details are available at https://www.decd.sa.gov.au/department/research-and-data/research-and-evaluation-department/conducting-research-and-evaluation.

Assisted by DECD, ACER will obtain Principal approval from schools participating intervention and control schools. Opt-in consent will be sought from the nominated teacher(s). The parents/carers of students in these nominated classes will be informed of their child's involvement in the study using an opt-out approach, but will not be told if their child is in an intervention or control class. Plain language statements will be prepared to explain the purpose and format of the evaluation for all participants – principals, teachers, students and parents/carers.

ACER are seeking confirmation from SVA regarding registration of the trial (e.g. By applying for an International Standard Randomised Controlled Trial Number (ISRCTN) at www.controlled-trials.com).

Risks

ACER General Policies

ACER has well-established procedures for minimising the risk of projects not achieving their objectives on time and within budget. All projects are designated an experienced Project Director who reports regularly to the relevant Head of Research Program. Dr Hilary Hollingsworth will carry out the role of Project Director, supported by Dr Katherine Dix as Senior Researcher. ACER has a highly-qualified staff of more than 390 people and is able to adequately replace any of the designated project team members if unexpected circumstances arise.

ACER has adopted a tailored PRINCE2© project management. The framework provides principles (manage by exception, learn from experience, defined roles and responsibilities, manage by stages, focus on products and tailoring), best practice and project life cycle steps (start up, initialisation, product delivery and closing) with an emphasis on dividing each stage into manageable, controlled phases. This methodology enables us to control timeframes, better manage risk and product quality, and control costs, whilst delivering the work package expected by SVA. We recognise that risk exists in all aspects of project operations, and take seriously the impact of risk on product delivery. Specific risk management techniques and standards are applied to ensure that risk is minimised in pursuit of meeting the objectives. This begins with flagging specific risks at the proposal stage, as provided in Table 2. The Project Director will create and maintain the Risk Register and ensure that risks are continually identified, assessed and controlled.

Risk management is an iterative process. Within each work package project and between projects, risk criteria will be strengthened to achieve progressively better levels of risk management. Identifying the risk involves considering what, why, when, where and how things happen, while analysing the risk involves developing an understanding of the risk, its consequences and the likelihood of the risk occurring. In evaluating each risk, we will determine those risks that are acceptable and those that require further treatment. To assess the level of risk, the likelihood of an incident occurring in combination with the seriousness of the consequence is analysed. The risk management responses shown in Table 4 explain how identified risks will be managed and outlines relevant contingency plans where applicable.

Table 4. Risk management matrix

Specific Risks	Risk Management Response
Insufficient interest from schools in participating	DECD have strategies in place to minimise this risk but the program and evaluation may have to be abandoned if sufficient schools do not engage. A minimum number of schools needs to be decided.
Insufficient number of participating schools due to staff	Allowances have been made in the calculation of sample size to comfortably achieve a detectible minimum effect size of 0.2. Multiple teachers per site increase the chance that each site will be represented. Paper-based feedback forms will be

Specific Risks	Risk Management Response
attrition and non-response	used at PL sessions to maximise response. A replacement teacher could continue the PL if they've only missed the first PL day.
Key ACER personnel unavailable	ACER has a large pool of qualified staff and is able to adequately replace any of the designated project team members if unexpected circumstances arise.
Key DECD personnel unavailable due to leave or illness	There has been no precedence of the program being delivered by other presenters. DECD have two presenters that work together. Sessions may need to be postponed or other personnel trained.
Insufficient DECD staffing	If the scale of the program increases beyond the work capacity of the two core presenters to the extent that new presenters are required, an effectiveness trial will need to be considered post-2017 in order to evaluate the effects of the program under more real-world conditions (Flay et al., 2005; Greenberg, 2004).
Capability of schools' computer connectivity	ACER will work with DECD to ensure that teachers can access the online surveys and avoid/respond to any firewall or other connection issues.
Analysis and reporting	All analysis executions, checking operations and planning of the report are planned in advance. All operations are closely monitored and reviewed by the Project Director.
Security of data	ACER's computer area has double-firewall security within an already secure environment. Our IT security systems use the latest technology from the top five security vendors and are constantly monitored and updated. Electronic access to the computer system is password protected. Only authorised project staff will have access to project data and files. All electronic and computer systems are backed up daily and a copy of materials kept in a secure off-site location. The security of the ACER computer system has been tested by an external agency and was reported to be highly secure.
Information Technology	ACER IT provides state-of-the-art IT infrastructure, communication, network and desktop support to the staff and customers of ACER. A Disaster Recovery site is in placed where all critical servers in the primary data centre are replicated to prevent functionality and data loss in an event of a disaster. Data will be synchronised in real time to the DR site.
Privacy and personal information	ACER regards the access to personal information for research purposes as an important privilege. Protecting the privacy of individuals whose personal information is used in our evaluations and the confidentiality of personal information in our custody (e.g. PATMaths data) is an integral commitment of ACER. For more information on our policies and procedures and ACER's statement on privacy protection, visit https://www.acer.edu.au/privacy
General	As issues arise the Project Director will maintain a Risk Log that highlights any perceived risks and documents the agreed resolution. SVA will be notified of any risks.

Personnel

ACER Evaluation Team

ACER will provide a team with extensive experience in evaluation and data analysis, mathematics education, and the professional development of mathematics teachers.

Dr Hilary Hollingsworth | Thinking Maths Evaluation Director, ACER
PhD; BEd; DipT (Primary)

Dr Hilary Hollingsworth is a Principal Research Fellow in ACER's Policy, Research and Practice Division. Hilary has specialised expertise in teacher professional learning, classroom observation frameworks, the assessment of student learning and mathematics education and she has led large scale research and consultancy projects internationally and nationally. In this project she will lead the evaluation team, playing a key role in the design, analysis, reporting and quality assurance.

Dr Katherine Dix | Senior Researcher Fellow, ACER
PhD Education; MPhil Science; BEd Hons, BSc Hons

Dr Katherine Dix is a Senior Research Fellow in the Educational Research and Monitoring division at ACER, located in the Adelaide Office. She has extensive experience in questionnaire design, administration and analysis (online and paper-based), and in quantitative research and evaluation methods, particularly in the use of multivariate and multilevel modelling. Katherine has expertise in assessing whole-school programs that focus on teacher capacity building and her honours thesis investigated innovative teaching practices in middle-school mathematics. Her innovative use of multivariate and multilevel data modelling techniques reported in the national evaluations of KidsMatter Primary and Early Childhood. She brings diverse experience in project managing and evaluating national school-based initiatives with a full understanding of moving an intervention from efficacy trials into national rollout. She will undertake the design, administration, analysis and reporting of the evaluation.

Dr Petra Lietz | Expert Adviser, ACER
BEd; MEd; CTEFLA; MAcc; PhD

Dr Petra Lietz is Head of ACER's Adelaide Office and Principal Research Fellow. Petra has specialised expertise in comparative research, secondary data analysis, statistics, and research design, augmented by her consultancy roles for the OECD and IEA, and her involvement with TIMSS, TALIS and PIRLS. Petra has also undertaken a number of evaluations and analyses for the South Australian Department for Education and Child Development (DECD). Most recently, she undertook the Australian Children Wellbeing Project (ACWP). Petra will provide local leadership and project management expertise in an advisory capacity.

Professor Kathryn Moyle | Expert Adviser, ACER
PhD; MEd; BEd (K-12)

Professor Moyle is a Research Director of Education Policy and Practice at ACER and has been in this role since 2014. She is responsible for the following portfolios: Indigenous

Education, Teaching and learning in School Education, Tertiary Education (Vocational Education and Training (VET) and higher education) and Program Evaluation. Her knowledge in these areas ensures she is familiar with the Australian Curriculum and Assessment Outline and various other State and International curricula. Professor Moyle has extensive experience working with the Australian Government and State and Territory governments in Australia and with the national Indonesian Government. She was also an invited member of Australian Curriculum Assessment and Reporting Authority (ACARA) Technology working party. Kathryn is the author of the Australian Education Review – Building Innovation: learning with technologies and has written and presented extensively about Information and Communications Technology (ICT) in school education. Her involvement in ICT in schools includes being E-Schooling manager for South Australian Department of Education, providing consultancy service to Education Services and Australia and education.au. She will provide advice through the ACER internal Expert Advisory Group for RCTs.

Dr Sheldon Rothman | Expert Adviser, ACER

BA; MAT; MEd Hons; EdD

Dr Sheldon Rothman is a Principal Research Fellow in ACER's Policy, Research and Practice Division. Sheldon has a wide range of experience in data analysis and the use of data in evaluations. He has provided advice to the Australian Government Department of Education and Training on the uses of NAPLAN data to set targets for student improvement under the Literacy and Numeracy National Partnerships and has managed major research projects. Sheldon will oversee the evaluation data collection and analysis process and provide input into the report.

Dr John Ainley | Expert Adviser, ACER

PhD; BSc

Dr Ainley is a Principal Research Fellow at ACER and retired Deputy CEO (Researcher) and Research Director of National and International Surveys Program ACER. He is a member of the Education and Training Statistics Advisory Group of the Australian Bureau of Statistics, the Consortium Advisory Group for the Longitudinal Study of Australian Children and the Youth Advisory Group for the Australian Institute of Health and Welfare. He is editor of the Australian Journal of Education and a member of the editorial boards of Education Research and Evaluation and the Education Research Review. He was a member of a group that conducted a national study of the impact of educational research. He will provide advice through the ACER internal Expert Advisory Group for RCTs.

Dr Siek Toon Khoo | Expert Adviser, ACER

BSc; DipEd; GrdDipCompSc; MEdSt; PhD

Dr Toon is Research Director of Psychometrics and Methodology at ACER. She is a measurement and modelling expert in the areas of multilevel modelling, causal modelling, longitudinal analysis, research design, assessment and psychometrics. She will provide advice through the ACER internal Expert Advisory Group for RCTs.

Juliet Young-Thornton | Administration, ACER
Juliet Young-Thornton is ACER's Adelaide Administration Officer. She provides office management and support for the Adelaide team. With a strong operational background, Juliet will provide project administrative support for the evaluation.

DECD Thinking Maths Program Team

Mr Ken Lountain | Thinking Maths Project Manager, DECD
Mr Lountain is Executive Leader, Strategic Design in the Department for Education and Child Development's Learning Improvement Division at DECD. He is experienced in the development, coordination and evaluation of a range of successful state-wide projects designed to improve learning outcomes in numeracy, literacy and across the curriculum. Mr Lountain will coordinate the project management of Thinking Maths.

Dr Pauline Carter | Thinking Maths Facilitator, DECD
BEd; MEd; PhD
Dr Carter is Project Officer, Critical and Creative Thinking Strategic Division, Learning Improvement Division, Office of Education and Early Childhood at DECD. An experienced secondary school teacher and Head of Mathematics in SA schools both public and private since 1978, Pauline completed a MEd (1996) and PhD (2008). Pauline has worked with the SACE Board in range of capacities. She has been involved in consultancy for the writing of the Australian Curriculum with ACARA and taught in Education at all three SA Universities. She has had a long association with the Mathematical Association (MASA), particularly in student activities, working as a fulltime Professional Officer at MASA from 2011-14. The recipient of numerous teacher excellence awards, Pauline was a Eureka Prizes Finalist (2011). She is co-presenter for the Thinking Maths intervention.

Ms Maureen Hegarty | Thinking Maths Facilitator, DECD
BEd; Grad Dip; MEd (Catholic Studies)
Maureen Hegarty is an experienced primary mathematics teacher. An educator for 26 years, she has taught in Catholic Education as a classroom teacher, key numeracy teacher and numeracy consultant (F-10). Maureen has lectured at UniSA and Flinders University in Maths Education and worked as Professional Officer at MASA. Maureen is currently Key Numeracy Teacher at Our Lady of Hope, Greenwith, SA and running professional development for teachers F-10. She codeveloped and copresented the DECD Teach SA and Year 6/7 Mathematical Pedagogy Knowledge and AC programs. She is co-presenter for the Thinking Maths intervention.

Ms Kathleen Ireland | Thinking Maths Assistant, DECD
Kath Ireland is a Project Officer in the Primary Learners Directorate in the Department for Education and Child Development. She is experienced in designing learning in the Australian Curriculum Mathematics and has managed state-wide initiatives to improve teaching and learning in Mathematics. Kath Ireland will manage the resource for Thinking Maths.

Timeline

The timeline below includes the Final Report to be delivered 30 April 2018 (assuming prompt access to 2017 PATMaths data for analysis). Progress Reports to SVA and DECD throughout the evaluation period are also proposed as contract deliverables.

Date	Activity	Responsible
22/6/16	2nd setup meeting	SVA
24/6/16	ACER Ethics Committee application submitted (Approved 30/6/2016, Ref no. 544883)	ACER
27/6/16	Thinking Maths Pilot PL Day 1: Grp1 27/6, Grp2 28/6 – ACER observe one	DECD
7/7/16	ACER Ethics Application/Letter to DECD (DECD to advise of process ASAP)	ACER/DECD
7/7/16	Submit Draft Trial Protocol Template – E4L to review during July	ACER/E4L
29/7/16	MOU between DECD and E4L provided for signing	E4L
29/7/16	Contract for ACER provided for signing	E4L
29/7/16	Thinking Maths Pilot PL Day 2: Grp1 29/7, Grp2 5/8 – ACER observe one	DECD
19/8/16	Thinking Maths Pilot PL Day 3: Grp1 19/8, Grp2 22/8 – ACER observe one	DECD
31/8/16	All Contract documentation finalised and executed	ACER/DECD
7/9/16	PATMaths student achievement (September- Weeks 7-10 Term 3) pre-test	DECD
12/9/16	Thinking Maths Pilot PL Day 4: Grp1 12/9, Grp2 16/9 – ACER observe one	DECD
12/9/16	School recruitment – EOI launched Week 8	DECD
24/10/16	EOI deadline – schools and nominated teachers (Week 2-3 Term 4)	DECD
31/10/16	Develop process evaluation and measurement tools (Term 4)	ACER
11/11/16	Thinking Maths Pilot PL Day 5: Grp1 11/11, Grp2 14/11 – ACER observe one	DECD
5/12/16	Consent to participate signed by selected school principals and Briefing invite sent to nominated teachers and principals (Week 8)	DECD

Date	Activity	Responsible
15/12/16	Progress Report 1 submitted to SVA – Learnings from Pilot Study, final evaluation tools/processes, a list of schools that have submitted signed consent to participate	ACER
6/2/17	Finalise and send nominated teachers and class lists to ACER	DECD
13/2/17	Recruitment Briefing Event(s) for evaluation participants (Week 3 Term 1)	ACER/DECD
13/2/17	Teacher and student pre-survey commence online (day of Briefing)	ACER
3/3/17	Teacher pre-survey closes	ACER
3/3/17	Randomisation of schools into control and treatment groups undertaken	ACER
6/3/17	Notification to teachers regarding placement in Intervention A (treatment) or B (control)	DECD
17/3/17	Student pre-survey closes	ACER
20/3/17	Data processing and analysis of pre-surveys commences	ACER
21/3/17	PATMaths 2016 data of participating students to ACER	DECD/ACER
20/3/17	Thinking Maths Intervention A commences – PL Day 1 sessions Week 8/9 Term 1: Grp1 8 Mon 20/3; Grp2 Fri 24/3; Grp3 Mon 27/3; Grp4 Fri 31/3	DECD
3/4/17	PL Day1 Feedback forms to ACER	DECD
8/5/17	Thinking Maths PL Day 2 sessions Week 2/3 Term 2: Grp1 Mon 8/5; Grp2 Fri 12/5; Grp3 Mon 15/5; Grp4 Fri 19/5	DECD
22/5/17	PL Day 2 Feedback forms to ACER	DECD
5/6/17	Thinking Maths PL Day 3 sessions Week 6/7 Term 2: Grp1 6 Mon 5/6; Grp2 Fri 9/6; Grp3 Tues 13/6; Grp4 Fri 16/6	DECD
19/7/17	PL Day3 Feedback forms to ACER	DECD
28/7/17	Progress Report 2 – Preliminary sample, pre-test, surveys and feedback results	ACER
28/7/17	Thinking Maths PL Day 4 sessions Week 1/3 Term 3: Grp1 Fri 28/7; Grp2 Mon 31/7; Grp3 Fri 4/8; Grp4 Mon 7/8	DECD
14/8/17	PL Day4 Feedback forms to ACER	DECD

Date	Activity	Responsible
21/8/17	Thinking Maths PL Day 5 sessions Week 5/6 Term 3: Grp1 Mon 21/8; Grp2 Fri 25/8; Grp3 Mon 28/8; Grp4 Fri 1/9	DECD
4/9/17	PL Day5 Feedback forms to ACER	DECD
4/9/17	PATMaths student achievement (September- Weeks 7-10 Term 3) post-test	DECS
18/9/17	Teacher and student post-surveys commence online	ACER
29/9/17	Teacher and student post-surveys close	ACER
3/10/17	PATMaths 2017 data to ACER	DECD/ACER
3/10/17	Data processing and analysis of post-surveys commences	ACER
9/10/17	Final data preparation and analyses (Term 4)	ACER
24/11/17	Progress Report 3 – Preliminary post-test, survey results, final sample stats	ACER
24/11/17	Draft Class Report template for review by DECD & E4L	ACER
1/12/17	Feedback on Draft Class Report template to ACER	DECD/E4L
11/12/17	Class Reports sent to participating teachers	ACER
16/3/18	Draft Final Evaluation Report to E4L for review	ACER
29/3/18	Draft Final Evaluation Report feedback from E4L	E4L
30/4/18	Final Evaluation Report	ACER

References

Arbuckle, J.L. (2007). *Amos User's Guide*. Chicago: SPSS.

Belfield, C., Levin, H.M., Hollands, F., Bowden, B., Cheng, H., Shand, R., Pan, Y., & Hanisch-Cerda, B. (2013). Cost-effectiveness in school education. Institute of Education Sciences Award Number R305U120001. http://www.alliance21.org.au/site/assets/media/Belfield_Alliance-21-Education-Innovation.pdf

Bloom, H.S., Richburg-Hayes, L. & Rebeck Black, A. (2007). Using covariates to improve precision for studies that randomise schools to evaluate educational intervention. *Education Evaluation and Policy Analysis*, 29 (1), 30–59.

Boaler, J. & Humphreys, C. (2005). *Connecting Mathematical Ideas: Middle School Cases of Teaching & Learning*. Heinemann: Portsmouth.

Boaler, J. (2002). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching and their Impact on Student Learning*. Lawrence Erlbaum Associates: Mahwah, New Jersey.

Boaler, J. (2015). *The Elephant in the Classroom. Helping Children Learn & Love Maths*. (2nd ed.) London: Souvenir Press.

Bryk, A.S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Newbury Park, California, Sage.

Chambers, J. & Parrish, T. (1994). Developing a resource cost database. In W. S. Barnett (Ed.), *Cost Analysis for Education Decisions: Methods and Examples* (Volume 4) (pp. 23-44). Greenwich, CT: JAI.

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

DECD (2016a). Pre-Easter response to SVA. DECD proposal communication to SVA.

DECD (2016b). Additional for SVA Yr7_8 150216. DECD proposal communication to SVA.

DECS (2010). *South Australian Teaching for Effective Learning (SA TfEL) Framework Guide*. Adelaide: South Australian Department of Education and Children's Services. learningtolearn.sa.edu.au/tfel/files/links/decs_sa_tfel_framework_gu_3.pdf

Dix, K.L., Slee, P.T., Lawson, M.J. & Keeves, J.P. (2012). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health*, 17(1), 45–51.

Domitrovich, C.E., Bradshaw, C.P., Poduska, J.M., Hoagwood, K., Buckley, J.A., Olin, S., et al. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, 1, 6–28.

Durlak, J.A. & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.

Dweck, C.S. (2000). *Self-theories: Their Role in Motivation, Personality and Development*. Philadelphia: Psychology Press.

Dweck, C.S., Walton, G. & Cohen, G. (2014). *Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning.* Bill & Melinda Gates Foundation.

ACER (2011). Interpreting ACER Test Results. Melbourne: Australian Council for Educational Research. https://www.acer.edu.au/files/PATM-Interpreting-Scores.pdf

Estimation 180 (2016). Estimation 180: Building number sense one day at a time. www.estimation180.com/

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151–175.

Greenberg, M.T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science*, 5(1), 5–13.

Hedges, L.V. & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Education Evaluation and Policy Analysis*, 29 (1), 60-87.

Hill, C.J, Bloom, H.S., Black, A.R. & Lipsey, M.W. (2007). *Empirical Benchmarks for Interpreting Effect Sizes in Research*. MDRC Working Papers on Research Methodology. New York, N.Y.: MDRC. www.mdrc.org/publications/459/full.pdf

Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S.E., Dixon-Woods, M., McCulloch, P., Wyatt, J.C., Chan, A.W. & Michie S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *The BMJ,* 348:g1687. www.bmj.com/content/348/bmj.g1687

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kirstin Kerr (2016). *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook*. Education Endowment Fund, University of Manchester, UK.

Hummel-Rossi, B. & Ashdown, J. (2002). The State of Cost-Benefit and Cost-Effectiveness Analyses in Education. *Review of Educational Research*, 72 (1), 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, H.M. (1995). Cost-effectiveness Analysis. In M. Carnoy (Ed.) *International Encyclopaedia of Economics of Education, 2nd Edition* (pp.381-386) Oxford: Pergamon.

Levin, H.M. & Garcia, E. (2012). Cost–effectiveness of accelerated study in Associate Programs (ASAP) of the City University of New York (CUNY). Retrieved from City University of New York. http://www.cuny.edu/academics/programs/notable/asap/Levin_Report_WEB.pdf

Levin, H.M., Belfield, C.R., Hollands, F., Bowden, B., Cheng, H., Shand, R., Pan, Y-L. & Hanisch-Cerda, B. (2012). *The Cost-Effectiveness of High School Completion Programs*. Monograph, Center for Benefit-Cost Studies in Education. www.cbcse.org.

Meyer, D. (2016). Dy/dan. Blog. http://blog.mrmeyer.com/

Miller, G. (2009). Influences on Student Learning by John Hattie. Summary by Gerry Miller. growthmindseteaz.org/johnhattie.html

Muthén, L.K. & Muthén, B.O. (2007). *MPlus User's Guide* (5th edn.). Los Angeles, CA: Muthén and Muthen.

nRich (2016). NRICH enriching mathematics. http://nrich.maths.org/

Radhakrishna, R. B., & Relando, R. Z. (2009). A framework to link evaluation questions to program outcomes. *Journal of Extension*, 47(3) Article 3TOT2. http://www.joe.org/joe/2009june/tt2.php

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, California: Sage.

Redmond, G., Skattebol, J., Lietz, P. et al (2016). *Are the kids alright? Young Australians in their middle years: Final summary report of the Australian Child Wellbeing Project*. Flinders University, UNSW Australia, Australian Council for Educational Research. www.australianchildwellbeing.com.au

Rossi, P.H. & Freeman, H.E. (1985). *Evaluation: A Systematic Approach*, 3rd Edition. Beverly Hills, California: Sage.

Scootle (2016) Scootle Community. www.scootle.edu.au

Sullivan, P. (2011). *Teaching Mathematics: Using Research-Informed Strategies.* Camberwell Victoria: ACER Press.

Sullivan, P.A., Clarke, D.M. & Clarke, B.A. (2013). *Teaching with Tasks for Effective Mathematics Learning.* New York NY USA: Springer.

SVA (2016) Innovative education grants program launches, May 24 2016. www.socialventures.com.au/news/evidence-learning-launches/

Timperley, H.W. (2008). *Teacher Professional Learning and Development. Educational Practices Series – 18*, International Bureau of Education. International Academy of Education.

Timperley, H., Wilson, A., Barrar H. & Fung I. (2007). *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration.* Wellington, New Zealand: Ministry of Education. www.oecd.org/edu/school/48727127.pdf

Torgerson, C.J. & D.J. Torgerson (2013). *Randomised trials in education: An introductory handbook*. London, Education Endowment Foundation.

Van de Walle, J. A. (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/