EVIDENCE FOR LEARNING

QuickSmart Numeracy

Learning Impact Fund Evaluation Report

A numeracy intervention delivered by Teaching Assistants to improve maths achievement in Primary and Secondary schools

Evaluation Report and Executive Summary

April 2019

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Independent Evaluators

About the evaluator

The project was independently evaluated by a team from the Teachers and Teaching Research Centre in the School of Education at the University of Newcastle, Australia:

- Dr Andrew Miller
- Laureate Professor Jenny Gore
- Dr Jess Harris
- Dr Elena Prieto-Rodriguez
- Dr Leanne Fray
- Dr Adam Lloyd
- Wendy Taggart

Contact details:

Andrew Miller
Teachers and Teaching Research Centre
School of Education
University of Newcastle, Australia
University Drive, Callaghan, NSW, 2308
Ph. 02 434 84214
Email. Andrew.miller@newcastle.edu.au
Reference: Drew, A, Gore, J, Harris, J, Prieto-Rodriguez, E, Fray, L, Lloyd, A, & Taggart, W 2019, 'QuickSmart Numeracy: Learning Impact Fund Evaluation Report'. Independent report prepared by the University of Newcastle for Evidence for Learning. Available from: https://evidenceforlearning.org.au/lif/our-projects/quicksmart-numeracy/. [April 2019].

Acknowledgements

This evaluation was commissioned and funded by Evidence for Learning. The evaluation team from the Teachers and Teaching Research Centre wishes to thank John Bush, Dr Pauline Ho and Matthew Deeble from Evidence for Learning, for their support at different stages of this evaluation.

We would like to thank the many school communities involved in this research, and recognise the sustained cooperation and support from the principals, teachers, and students involved. We are also grateful for the support of Sydney Catholic Schools, and particularly Justin Slowey.

We would also like to recognise the work of Professor John Pegg and Anne Parnell from the SiMERR National Research Centre for their cooperation with the evaluation and for their ongoing commitment towards literacy and numeracy improvement in Australia.

Executive Summary

The project

QuickSmart Numeracy (‘QuickSmart’), developed by the SiMERR National Research Centre, is an intensive 30-week tutoring intervention which aims to increase fluency and automaticity in mathematics for students in Years 4 to 9 (aged between approximately 9 to 15 years) performing in the bottom third of their national cohort in mathematics (based on NAPLAN results).

Across the intervention period of 30 weeks, pairs of participating students are withdrawn from class for three 30-minute sessions per week, for a desired intervention total of 90 sessions. The intervention is delivered by a trained school staff member (QuickSmart instructor), typically a teaching assistant (‘TA’). Schools are also supplied with program user guides, lesson formats and content, instructional resources, and the Cognitive Aptitude Assessment System (OZCAAS) software for ongoing assessment throughout the duration of the QuickSmart program.

To prepare for the intervention and to support the ongoing delivery of QuickSmart, instructors attend three two-day workshops in their first year of delivering the intervention. Further training is available for instructors after their first year, with three optional one-day workshops in the second year, and a single optional one-day workshop in the third and subsequent years an instructor delivers the program. The level of training and delivery experience may vary between schools, and across years of delivery within a school (e.g., schools may have delivered the program for four years but are using first year instructors due to staff turnover), however the intervention delivered to students (90 sessions over 30 weeks) is the same in structure and format regardless of the level of instructor training or experience.

The evaluation of QuickSmart was independently conducted by the Teachers and Teaching Research Centre at the University of Newcastle between January 2017 and May 2018. Supported by Sydney Catholic Schools, this QuickSmart evaluation involved 288 Years 4 and 8 participating students from 12 Primary and 11 Secondary schools in the Diocese of Sydney. This evaluation was designed as an effectiveness trial seeking to obtain the effects of the QuickSmart intervention when delivered in its usual state across a group of schools (e.g., as delivered outside of an intervention). The decision to run a ‘real-world’ evaluation, rather than an efficacy trial, was based on the previous and wide-spread scaling of the QuickSmart program, with the relevant information regarded as “does it work as currently implemented in schools” rather than “does it work when implemented perfectly by the research team”.

This evaluation was a multi-site randomised controlled trial, involving randomisation at the individual level within class groups (an intervention and a control group within each class). The trial was designed to identify the impact of the QuickSmart intervention on mathematics achievement beyond the impact of regular classroom mathematics instruction provided by the classroom teacher, comparing an intervention and control condition matched using their baseline mathematics achievement scores in each class group. Whilst this design theoretically reduces bias at the teacher level by assuming the same quality/type of classroom instruction for students from both conditions, the teacher knowing the allocation of the participants may produce bias in the form of preferential attention during mathematics-based instruction. It is not possible to predict the direction of this bias (control of intervention preferential treatment), and the inclusion of many classes in this trial (70 classes total) potentially negates this effect, however it must be recognised as a limitation of this research design in determining the effectiveness of the QuickSmart Numeracy program.

The primary outcome of mathematics achievement was assessed using the Australian Council for Educational Research (ACER) Progressive Achievement Test – Mathematics (PAT-M), with secondary outcomes of mathematics self-efficacy, interest, self-concept and anxiety measured using instruments developed for the Programme for International Assessment (PISA). Baseline measures were taken in March 2017 prior to randomisation, with follow-up measures taken in May 2018, six-months after completion of the intervention period in December 2017. Interviews were undertaken with students from two primary and two secondary schools at eight-week intervals throughout the intervention to examine students' perceptions of QuickSmart and engagement with school and mathematics. Interviews with mathematics teachers and QuickSmart Instructors from this sub-sample provided deeper understanding of how the QuickSmart program was implemented in their schools.

In this trial, instructor experience and level of training varied across the sample due to the recruitment of new and existing QuickSmart schools into the evaluation, with some existing QuickSmart schools also using first year trained instructors for the delivery of the program. There were eight new schools recruited for this trial (35% of schools), with the majority of these in the Secondary cohort (7/11; 64%). It is recognised by the evaluation team that the use of either all first-year schools and instructors, or specification of some base level of school experience and instructor training (e.g., minimum of one year delivery experience) would make for clearer interpretation of results. As there is no mentoring structure or minimum demonstrable instruction standard in place within the implementation model of QuickSmart, instructors are deemed as proficient from the outset of their involvement, and the inclusion of new and existing schools/instructors sits within the model of an effectiveness trial as this is how QuickSmart is undertaken across school systems. There is no information recorded or reported by SiMERR regarding the training and experience of QuickSmart instructors, and whether the balance of training and experience was common in this trial is unknown.

Key Conclusions

1. In this trial, QuickSmart did not have an additional impact on maths achievement compared to regular classroom instruction and support. There was a small positive gain, equivalent to one month's additional learning, however this trial was not commissioned to detect this level of difference[1] meaning the difference was not statistically significant.

2. When models were adjusted for intervention exposure, there was a small increase in the effect on student achievement (indicating that exposure levels have some effect on outcomes), however this effect was not statistically significant.

3. Sub-group analysis displayed a small but not statistically significant positive effect for Primary students. The gain was equivalent to one month's additional learning. There was no additional effect for Secondary students.

4. Schools faced challenges achieving the prescribed program exposure of 90 sessions within 30 weeks. Primary students, on average, received 73% (or 66 sessions) of QuickSmart's prescribed 90 sessions over 30 weeks, while Secondary students received 49% (or 44 sessions). Only 35% of Primary students and 4% of Secondary students received more than 75% (or 67 sessions) of the prescribed QuickSmart sessions.

5. Sound implementation of QuickSmart appeared more feasible within Primary schools than Secondary schools. Both settings struggled with transitions into and out of the classroom, and concern about the subject matter students were missing out on as a result of QuickSmart was expressed across Primary and Secondary settings.

6. Primary teachers were positive about QuickSmart and reported that it appeared to help students gain more confidence participating in their maths classrooms. QuickSmart had a statistically significant positive impact on Primary students' maths self-concept (effect size g = 0.30) and interest in maths (effect size g = 0.47), however there was no evidence of impact on self-efficacy (effect size g = 0.09). There were no statistically significant intervention effects on Secondary students' cognitive and affective outcomes.

[1] This trial was powered to achieve a Minimum Detectable Effect Size (MDES) of 0.24 at randomisation, which meets the high padlock rating criteria for MDES of <0.3.

Main findings and impact

This trial did not produce significant evidence that participation in the QuickSmart program had a positive effect on the average mathematics achievement beyond participating in regular classroom-based mathematics instruction.

In this trial, intervention and control groups received the same amount of time in which to demonstrate improvements, and there was no continuation of QuickSmart sessions beyond the single school year. Because of this important design feature for controlled trials (equal time among groups), and the trial processes of recruitment and testing, not all schools recruited had access to the 30-week intervention period required to undertake the complete intervention volume of 90 sessions. Of the 23 schools involved, 12 (52%) had access to 30-weeks of intervention time, eight (35%) had 28 – 29 weeks, two (9%) had 26 – 27 weeks, and one (4%) had only 23-weeks of intervention time. When comparing the average exposure to the program for this trial against exposure reported from 2013 – 2016 by SiMERR (Pegg et al., 2013, 2014, 2015, 2016), trial average exposure was marginally greater among the Year 4 cohort (Trial = 73%; SiMERR = 66%), and marginally lower among the Year 8 cohort (Trial = 49%; SiMERR = 55%).

When models were adjusted for intervention exposure, there was a small increase in the effect on student achievement, and there were also signs of stronger effects in the Primary school context (equivalent to one month of additional learning progress)[2], than observed among the Secondary schools. Whilst neither of these findings were statistically significant, they do indicate that exposure levels appear to have some effect on outcomes.

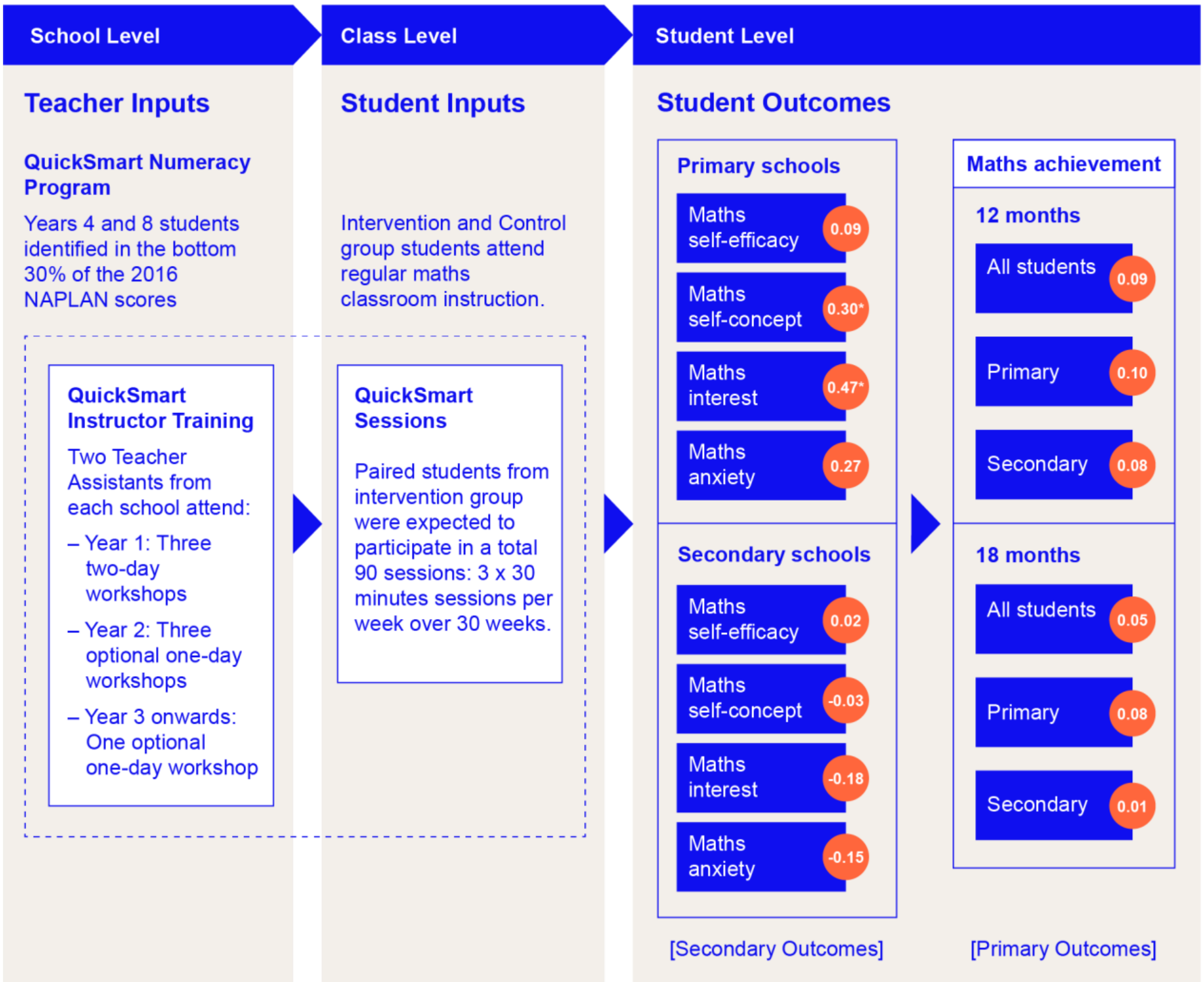
QuickSmart displayed positive effects on the cognitive and affective measures of mathematics self-concept and mathematics interest among the Primary cohort. The process evaluation highlighted that instructors, teachers and students valued the intervention, reporting a positive effect on students' confidence in mathematics, and adding support for the quantitative outcomes.

Schools reported disruption for students as they transitioned between the classroom and QuickSmart instruction as well as reduced learning time in other subject areas due to QuickSmart sessions. Primary schools in our evaluation appeared better able to implement the QuickSmart intervention, with students able to participate in a greater volume of QuickSmart sessions on average than the Secondary cohort. QuickSmart Instructors in Primary schools were, on average more qualified and had more years of QuickSmart experience, perhaps leading to increased efficiency within Primary schools. Stronger implementation support for schools and coordination with classroom teachers may assist with the smooth and effective delivery of QuickSmart and help achieve the greatest possible return on investment.

---

[2] Months progress is converted from the effect size using Evidence for Learning's conversion table (see Appendix A).

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



*indicates statistically significant effect (p<0.05)

Research results

Mathematics achievement	Effect size [95% CI]	Estimated months progress*	E4L security rating**	Number of students (Intervention, Control)	P value	E4L cost rating***
Mathematics achievement	0.05 (-0.19, 0.30)	+1	🔒🔒🔒🔒	287 (145, 142)	0.59	$$$$$ ($$$)
Primary	0.08 (-0.28, 0.44)	+1	🔒🔒🔒🔒	133 (67, 66)	0.48	$$$$$ ($$$)
Secondary	0.01 (-0.33,0.35)	0	🔒🔒🔒🔒	154 (79, 76)	0.95	$$$$$ ($$$)

* 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.
*** When staffing costs are included, the cost rating for QuickSmart is Moderate.

Evidence for Learning's security rating – how secure is this finding?

Findings from this trial have a high degree of security with four padlock rating. The trial was set up as a randomised control trial, with student-level randomisation. This was an effectiveness trial, which sought to test whether the supplementary maths instruction can work at scale, under the everyday conditions of schools. Although there has been published evidence for the effectiveness of QuickSmart, there has been no evaluation to assess the effects of QuickSmart as it is currently undertaken within schools using a randomised control group from a homogeneous sample of students who are all QuickSmart eligible.

Through the recruitment process, a maximum of 152 students from 12 Primary schools (30 classes) and 169 students from 11 Secondary schools (40 classes) were randomly assigned to either an intervention group that received QuickSmart and or a control group that received only 'business-as-usual' classroom teaching. The key conclusion is that students who received QuickSmart made no additional months of impact compared to students who only received regular classroom teaching. There was an effect size benefit of 0.05, equivalent to one month's additional progress however this result was not statistically significant. The program effect was not large enough to be detected as this trial was not commissioned to detect this level of difference.

The trial had a well-balanced randomised sample (difference in pre-test of 0.02 SD in favour of the control group). The attrition was low; in total, results from 288 students were assessed compared to 304 students who began the trial (attrition of 10.5%). The attrition was mainly due to students moving school and sickness on the day of testing. There was no evidence that the attrition had biased the results and despite the attrition on the randomised sample, there was a good sample at baseline for the analysed sample (g= -0.02, p=0.89). The attrition and the MDES of this trial reduced the security of the findings by one padlock (Appendix B).

A limitation of the study is that TAs in schools had previously received training in the intervention prior to the study, while TAs in other schools were new to the QuickSmart approach. The varying levels of Instructor Training exposure prior to the study introduce the risk of variability in QuickSmart delivery as TAs who received more QuickSmart Instructor training before the study may be more proficient with the program than those who just started training.

Blind marking of test papers was undertaken, but classroom teachers, particularly Primary school teachers, knew which students were allocated to each of the study conditions. There is the possibility that students in the control or intervention group may receive more attention in class or given other forms of intensive support in the regular maths class. As classroom observations were not undertaken, it was not possible to determine the direction of this bias of any preferential treatment on either control or intervention group. However, given the inclusion of many classes in this trial (70 classes total), the potential for this to affect the performance of the intervention or control group results is considered minor.

How much does it cost?

The initial outlay for a school to undertake the QuickSmart program is $10,500 (exc. GST). This cost, payable at start-up includes access to six days of training in the first year (for up to 5 staff), access to online resources and telephone support ($7,000), with the remaining costs ($3,500) covering the required equipment and resources, and a three-year licence to the OZCAAS program required for delivery and assessment. If a school continues to run QuickSmart past three years, an additional three-year OZCAAS licence will cost $1,800.

The cost per student is estimated at $151 per year, based on 25 students per year undertaking the intervention, and the program continuing over a three-year period at a school. This estimate includes licensing and equipment costs for the program ($140 per student) and printing and stationery costs ($11 per student), but does not include costs associate with direct staffing, teacher release to attend training or equipment. When staffing costs are included in the cost analysis, the cost per student is estimated at $1,007 per year, based on 25 students per year undertaking the intervention, and the program continuing over a three-year period at a school.

The cost per student is rated as very low according to the Evidence for Learning Cost Rating approach when staffing costs are not included in the estimate, and moderate when staffing costs are included (see Appendix A).

Table of Contents

List of Tables

List of Figures

1 Introduction

This report presents the findings of the outcome and process evaluation of the SiMERR National Research Centre's QuickSmart Numeracy program, independently conducted by the Teachers and Teaching Research Centre at the University of Newcastle. This report is designed to be read by as wide an audience as possible. Further technical details of the defined study protocols can be found in the statistical analysis plan (SAP).[3]

1.1 Intervention

QuickSmart Numeracy is an intensive 30-week tutoring intervention for Primary and Secondary school students (grades 4-9 or 9-15 years of age) performing in the bottom third of their national cohort in mathematics. The primary aim of the program is to develop students' fluency and automaticity in mathematics. Pairs of students are withdrawn from class for three 30-minute lessons a week, with lessons usually delivered by Teaching Assistants (TAs) in a dedicated room within the school, with access to computers. To prepare for the intervention, teachers attend three two-day workshops in the first year a school runs QuickSmart, three optional one-day workshops in the second year, and an optional single one-day workshop for the third and subsequent years. They are also supplied with program user guides, lessons formats and content, instructional resources, and the Cognitive Aptitude Assessment System (OZCAAS) software for ongoing assessment throughout the duration of the QuickSmart program.

The approach is based on research showing that students with learning difficulties are slowed down by their lack of automaticity in recalling basic number facts. The intervention involves the use of guided and independent practice activities (such as flash cards and speed sheets) that develop students' ability to readily recall basic number facts. QuickSmart sessions also aim to develop students' understanding of key mathematical concepts e.g. their ability to identify and use number patterns and understanding of the meanings of and relationships between different operations.

[3] Evidence for Learning. E4L QuickSmart Numeracy Statistical Analysis Plan. January 2018.

1.2 Background evidence

According to the QuickSmart developers, students who experience ongoing failure in upper-Primary and lower-Secondary school face substantial difficulties in pursuing post-school options and contributing to society through employment and active citizenship (Pegg et al., 2013). Those students who exhibit consistent weaknesses in basic skills, such as the recall of number facts, are particularly vulnerable (Pegg & Graham, 2013).

The QuickSmart instructional approach focuses on the role of automaticity in developing students' fluency and facility with basic academic facts, and is informed by relevant literature associated with learning difficulties/disabilities and quality instruction (Baker, Gersten, & Lee, 2002; McMaster, Fuchs, Fuchs, & Compton, 2005; Westwood, 2007), effective instruction (Rowe, Stephanou, & Urbach, 2006), mathematics education (Fuchs & Fuchs, 2001) and educational interventions (Deshler, Mellard, Tollefson, & Byrd, 2005; Marston, 2005).

Since 2001, more than 1,300 schools from across the country have used the QuickSmart program. During that time period, SiMERR has evaluated QuickSmart using a quasi-experimental design, comparing QuickSmart intervention students selected for being in the lowest third of national NAPLAN results with a small sample of average-achieving students (mid-third of NAPLAN numeracy) in a comparison group. Results published in QuickSmart annual reports from 2013 - 2016 by the SiMERR team using this design have demonstrated average effect sizes among the QuickSmart group of d = 0.942 and d = 0.639 for Year-4 and Year-8 students respectively for the Progressive Achievement Test – Mathematics (Pegg et al., 2013, 2014, 2015, 2016). These results are compared to the average comparison group effects of d = 0.657 and d = 0.262 for Year-4 and Year-8 students respectively on the same measure across the same reporting time frame (2013 – 2016).

Whilst the results reported from 2013 – 2016 suggest an average effect size difference between groups of d = 0.305 and d = 0.377 for Year-4 and Year-8 students respectively, suggesting approximately 3-4 months additional achievement growth for the QuickSmart group, there are research design concerns to take into consideration. The quasi-experimental research design employed to evaluate the QuickSmart program does not employ randomisation of participants prior to the intervention, resulting in unequal treatment conditions, both in group size and baseline outcome measures. There is no indication from the reporting of the research if participants from the two conditions are assessed over the same time period, and pooled statistical analysis comparing the two conditions was not undertaken.

Given that no evaluation of the impact of the program on students' mathematics achievement has been conducted using a randomised control group drawn from a homogeneous sample who are all QuickSmart eligible, this evaluation was designed to assess the effects of the QuickSmart intervention using a randomised experimental design. The trial was designed as an effectiveness evaluation to assess the QuickSmart intervention as it is currently undertaken within schools due to the broad scaling that has already occurred for this intervention. In this case the question was not about the effect of the intervention under ideal conditions, but the effectiveness as is commonly delivered in schools.

1.3 What does the program involve?

The program is designed to run for 30 weeks, excluding a two-week setup and close-down period, with sessions typically overseen by TAs, and sometimes teachers or school executive members.

QuickSmart instructor training

QuickSmart instructors are required to attend three two-day workshops in the first year a school runs QuickSmart; three optional one-day workshops in the second year; and an optional single one-day workshop for the third and subsequent years. These workshops typically occur at a central location, with instructors from many schools attending.

The workshops attended in the first year are considered basic skills training in QuickSmart, which result in attendees being certified as QuickSmart Instructors. These workshops introduce the approach to participants, who learn about and discuss the program's underlying theoretical perspectives, trial the materials, refine their teaching and assessment techniques, and share their experiences with peers.

In the second and third workshops, teams from between 10 and 15 schools discuss any issues that they have faced in implementing the program. The workshops also emphasize issues related to providing information to parents about QuickSmart and encouraging their involvement in and support of the program in their child's school. Workshop participants share successful strategies for encouraging parent involvement.

The focus of the workshops undertaken in years two and three focus on advanced skills training. Here participants review the central ideas addressed in the first year in a deeper way. In particular, sessions focus on:

- the cognitive and neuroscientific underpinnings;
- features such as deliberate practice, formative assessment, and feedback;
- issues associated with learnable skills; and,
- evaluative frameworks that are suitable in a QuickSmart environment.

In summary, the professional learning program that accompanies the QuickSmart program is focused on supporting instructors to understand and provide:

- effective instruction that maximises student on-task time, and scaffolds learning to ensure that students experience improvement and success;
- a motivational environment that is safe for vulnerable learners and builds trust between Instructor and students;
- deliberate practice that is integral to every lesson, allows for success and is focused on providing targeted feedback to improve learning;
- guided and independent timed practice activities;
- strategy instruction and concept development;
- evidence of competence and increasing student confidence by encouraging a 'can do' and 'have a go' attitude;
- appropriate teacher and peer modelling; and,
- motivational academic activities that provide opportunities for modelling and for developing fluency.

Lesson Delivery and Resources

Back in their schools, instructors lead students through a specific set of activities, focused on particular numerical operations and focus-facts. The activities are designed to target students' current level of mathematical proficiency with those operations and facts. The overarching aim of QuickSmart is to increase students' accuracy and automaticity of basic numeracy skills (Pegg & Graham, 2013). Specifically, the pedagogical approach used focuses on a variety of practice and recall strategies geared to developing understanding and fluency of basic numeracy skills. Each lesson involves revision of the previous session, a number of guided and deliberate practice activities featuring overt self-talk, discussion and practice of memory and retrieval strategies, timed speed sheet activities followed by independent practice activities, and an educational game.

Ongoing assessment and instruction form a continuous cycle. The instruction delivered by one adult to two students is personal, connected and targeted. Instructors' observations and the information gained from questioning students about their knowledge and strategy use form the basis of instructional decision-making and individualisation. Assessment and lesson delivery is supported by a package of teaching and learning resources, including a teacher manual, a resource folder, flash cards, a games pack and a license for the Cognitive Aptitude Assessment System (OZCAAS) program (for additional details, refer to: https://simerr.une.edu.au/quicksmart/pages/qsprogram-materials.php). OZCAAS works on a computer program that provides time and accuracy scores on randomly generated basic mathematics operations. QuickSmart licensing also provides schools with access to a QuickSmart website which contains a range of teaching materials, many developed by QuickSmart schools.

The timed performance activities use the OZCAAS program, which provides ongoing data to students and the instructor about the student's improvement in speed and accuracy. In the last part of the lesson, students practise their skills independently on worksheets that are closely related to the lesson content.

Data and monitoring student progress

Before the 30-week intervention, there is a two-week set-up phase. In this set-up phase, students are introduced to the program and their role. During this time, delivery staff use OZCAAS to collect pre-intervention data on the speed and accuracy on basic numeric skills, as well as performance on an independent test (of age-appropriate material but not mathematics topics specifically part of the intervention). These data provide baseline information to guide the instructor on where instruction might most profitably begin. Additional data is collected in each lesson to give students and the instructor information about the student's improvement in speed and accuracy in relation to basic mathematics facts. Following the 30 weeks of instruction, students are re-tested to obtain post-intervention data. The style and substance of the testing is equivalent to what students have experienced previously as part of their QuickSmart lessons.

Program fidelity and implementation

Schools are responsible for ensuring the program is delivered in the manner intended by the designers. Schools are required to report the student outcomes data (pre and post program), and the volume of program exposure (total number of sessions undertaken) back to the SiMERR National Research Centre, however there are no external checks regarding the quality of the program sessions, or the frequency with which sessions are undertaken.

1.4 Control condition

In this evaluation, the control condition was 'business as usual', with students attending regular mathematics and other classes. Students in the control condition were offered the intervention following the completion of the evaluation (wait-list control).

1.5 Evaluation objectives

This mixed methods evaluation offered a systematic and comprehensive approach to understanding the effects of the QuickSmart intervention. The primary intervention outcome was student achievement in mathematics. Secondary outcomes included student self-efficacy, interest, self-concept and anxiety. QuickSmart was evaluated using a multi-site two-arm individual randomised trial comparing intervention and control condition students from within the same class group (maths class group for Secondary cohort students). This design was chosen in an attempt to control potential bias produced at the class level by the influence of the classroom teacher when randomisation occurs at the class level. Whilst this design theoretically provides the intervention and control participants with the same type of instruction during class based activities, the teacher knowing the allocation of the participants may produce bias in the form of preferential attention during mathematics based instruction. This bias may not necessarily favour the control group however, as a teacher may believe in the intervention as a positive and give preferential treatment to the intervention group. With the inclusion of many classes in this trial (70 classes total), we believe that the potential for bias from teacher preferential treatment is negated, however it must be recognised as a limitation of this research design in determining the effectiveness of the QuickSmart Numeracy program.

Research questions

The specific research questions for this evaluation were as follows:

1. What is the impact of QuickSmart on student mathematics achievement?
2. What is the impact of QuickSmart on student self-efficacy and engagement in relation to mathematics?
3. What is the impact of QuickSmart on student experience within the classroom and the broader school context?
4. What is the cost per student of the QuickSmart program?

1.6 Ethical review

Ethical reviews were completed by the University of Newcastle's Human Research Ethics Committee and Sydney Catholic Schools (SCS) prior to school recruitment (approval number H-2016-0338). Following recruitment, the evaluation team obtained active informed consent to be involved in the evaluation from all participating schools, QuickSmart instructors, teachers and students (in conjunction with their parents/carers).

Hard

1.7 Project team

The evaluation team was led by Dr Andrew Miller of the Teachers and Teaching Research Centre, University of Newcastle (UON). Roles of the evaluation team members were as follows:

- Laureate Professor Jenny Gore – oversight of the evaluation;
- Dr Andrew Miller – management of RCT and quantitative data;
- Dr Jess Harris – management of qualitative data;
- Dr Elena Prieto-Rodriguez – management of efficacy data;
- Dr Leanne Fray and Dr Adam Lloyd – consultancy on the evaluation; and
- Wendy Taggart – project management and coordination of data collection.

The QuickSmart intervention was led by Professor John Pegg, of SiMERR National Research Centre, University of New England, with QuickSmart Research Fellow Anne Parnell delivering and overseeing the intervention in the evaluation schools.

1.8 Trial registration

The randomised controlled trial conducted as part of this evaluation was registered in November 2016 with the Australian New Zealand Clinical Trials Registry (Trial ID no. ACTRN12616001612404).

2 Methods

2.1 Trial design and overview

QuickSmart was evaluated using a multi-site two-arm individual randomised trial comparing intervention and control condition students from within the same class group. Intervention students within a class received the QuickSmart program in addition to regular classroom tuition, while the wait-list control condition students in the same class received just their regular classroom tuition. This design provided an intervention and control condition within each class group, accounting for the effect of the classroom teacher on students in both conditions, and enabling evaluation of the effect of the QuickSmart intervention in addition to regular classroom mathematics instruction.

Primary (Year 4) and Secondary (Year 8) school cohorts were recruited for the evaluation. Students identified in the bottom 30% of the most recent NAPLAN round (Year 3 for Year 4 cohort, Year 7 for Year 8 cohort), with no existing diagnosis of a learning disorder were eligible to participate. The recruitment target was 24 schools (12 Primary / 12 Secondary). This design was based on an approximation of 12 students from each Primary school, and 16 students from each Secondary school (regardless of the number of classes). In total, 23 schools were recruited, resulting in 146 intervention students and 142 wait-list control students (n = 288) from 70 class groups after randomisation. Details of each cohort are recorded below:

Primary: 12 schools; 30 classes; 67 intervention / 66 control (n = 133)

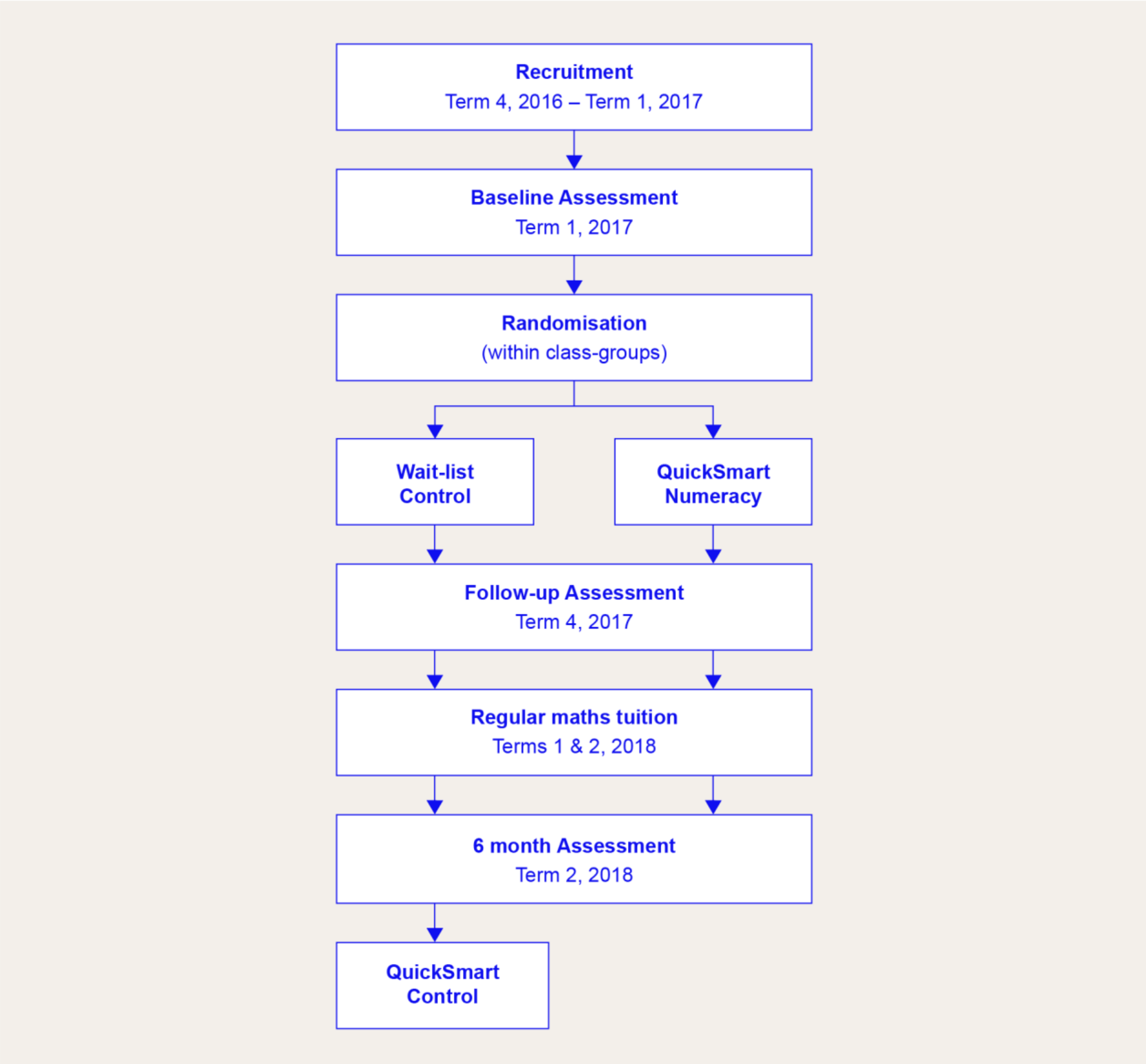
Secondary: 11 schools; 40 classes; 79 intervention / 76 control (n = 155)

All participating students completed baseline measures prior to randomisation during Term 1 (March), 2017. Students in both conditions received regular classroom mathematics tuition throughout 2017, and students allocated to the intervention condition participated in the QuickSmart program for the 2017 academic year. Follow-up measures were collected at the end of Term 4 (November), 2017, and again in Term 2 (May), 2018 at 6-months post intervention.

As QuickSmart students received more frequent numeracy instruction than those in the control group in 2017, the 6-month follow-up assessment (reported here) was used to assess the effects beyond immediate exposure to the QuickSmart intervention. Results from the immediate follow-up time point (November 2017) are reported as an interim analysis in the results below for the primary outcome to give an indication of the immediate intervention effects. Students allocated to the wait-list control group were offered the opportunity to participate in the QuickSmart program from Term 2, 2018, after all intervention data was collected.

Note. Reporting of the immediate follow-up time point (November 2017) was not specified in the original protocol.

Figure 1. Evaluation Design



2.2 Participant selection

Recruitment

Students identified in the bottom 30% of the most recent round of National Assessment Program - Literacy and Numeracy (NAPLAN), with no existing diagnosis of a learning disability were invited to participate by the University of Newcastle evaluation team via the following process:

1. Sydney Catholic Schools (SCS) identified the number of eligible students enrolled in each school in their diocese for 2016, and shared a list of schools with their respective number of eligible students (no student identification) with the University of Newcastle (UON) and University of New England (UNE) research teams.
2. Schools with at or above the desired numbers of students per school (Primary = 12, Secondary = 16) were selected to ensure appropriate student numbers to maintain clustering, sample size and QuickSmart program logistics (pairing of students and appropriate numbers for the staff load required). These schools were invited to a presentation on the program and evaluation, with SCS, UNE, UON and Social Ventures Australia (SVA) representatives present.
3. Schools were invited by UON, and upon giving consent were asked to distribute recruitment documents addressed to parents for the students identified by SCS.

Consent to participate

The Principal (or their delegate) of all participating schools provided their informed written consent to participation in the evaluation to the evaluation team. Following this, informed written consent to be involved in data collection activities was provided to the evaluation team by all participating QuickSmart instructors, teachers and students (in conjunction with their parents/carers). Consent from all participants was provided prior to their initial assessment.

2.3 Outcome measures

Assessment items were administered by the University of Newcastle researchers at all time-points (See Table 2. Study Timeline). The researchers involved in administering and scoring the assessments were blinded to student allocation to intervention or control groups.

Primary outcome

Progressive Achievement Test – Mathematics (PAT-M)

The primary intervention outcome was measured with the Australian Council for Educational Research (ACER) Progressive Achievement Test – Mathematics (PAT-M). The PAT-M is a rigorously tested measure of mathematics achievement that is well suited for evaluation of this project because each year level test is designed to be developmentally appropriate (e.g., Year 4; Year 8). Administered and scored using a paper-based format, participants completed the same level PAT-M at each assessment time-point (Year 4: PAT-M 3rd ed. Test Booklet 2; Year 8: PAT-M 3rd ed. Test Booklet 5).

The test scaled score was used for analysis. Scale scores are measures on an interval scale (0 to 100). This means that a difference of five in scale scores in the middle of the PAT scale (for example, from 50 to 55) is equivalent to the same difference on any other part of the scale (for example, from 15 to 20 or from 85 to 90). Scale scores allow comparison of results on test booklets of varying difficulty, and provide a common achievement scale for all test booklets.

Secondary outcomes

Student cognitive and affective measures

Mathematics self-beliefs have an impact on learning and academic achievement as they determine how well students are able to motivate themselves and persevere in the face of difficulties. They influence students' emotional life, and they affect the choices students make about their educational and career paths (Bandura, 1997; Wigfield and Eccles, 2000).

Self-beliefs were measured using instruments developed for the Programme for International Assessment (PISA). These instruments were used without modification with the Year 8 cohort and with slight changes in wording for the Year 4 cohort to make the language used accessible to young students. Administered in paper-based format, the four PISA scales used were:

- Mathematics self-efficacy scale (MATHEFF), comprised of eight items;
- Interest in mathematics scale (INTMAT), consisting of four items;
- Self-concept scale (SCMAT), containing five items; and
- Mathematics anxiety scale (ANXIMAT), also containing five items.

Items were rated on a 4-point Likert scale ranging from (1) "Strongly agree" to (4) "Strongly disagree", with the average of the items used for analysis.

The 2003 and 2012 PISA technical reports, and research from which the questions and calculation of measurement scales were based, highlight both valid and internally consistent scores (Wigfield et al., 1997; Ferla, Valcke & Cai, 2009).

2.4 Sample size

The calculated sample size included considerations of statistical power and access to a convenient sample indicated by the intervention research team. Sampling was calculated on a per-cohort basis (Primary and Secondary) to ensure sufficient recruitment for analysis to take place for each cohort. The sample size calculation was based on a three-step process:

1. Calculation for a linear approach to modelling (repeat measures ANCOVA)

G*Power (version 3.1.9.2) was used to determine an unadjusted sample for the desired effect, type 1 error and power. An ANOVA F test (Repeated measures, within-between interaction) was used as the correlation among repeated measures could be taken into consideration during calculation. Assumptions: Effect *f* = 0.15, Alpha = 0.05, Correlation among measures = 0.5, Power = .80

The estimated effect of 0.15 is considered small-to-medium within an F-test (Cohen, 1992). As the lowest 30% of NAPLAN achievement were recruited, a low correlation among repeated measures was used as a low pre-post correlation was expected due to the lack of variance available in this group. A sample of 90 students was required for the unadjusted (for clustering) cohort.

2. Adjustment for clustering

Clustering in educational studies has the effect of reducing the amount of data available during analysis (Dreyhaupt, Mayer, Keis, Öchsner, & Muche, 2017). The more students with groups (e.g., within schools) are like each other, and the more their groups differ from each other (e.g, between schools), the closer each group moves to acting like a single data point during analysis, reducing the power available. To adjust for clustering, the correction factor [1+ (m – 1) x ICC] was applied (Donner & Klar, 2000), where m = students per school and ICC = the intra-class correlation coefficient (between school variance / (between school variance + within school variance). Assumptions: rho(w) = 0.05 (correlation coefficient for within cluster variation), subjects per cluster (school) = 10.

The PISA 2012 Technical Report suggests an Australian ICC for mathematics of 0.28 (OECD 2014, p.439). However, based on data from Lamb & Fullarton (2001) demonstrating between school variance of 0.104 when between class variance of 0.279 is taken into account at the second level of analysis (resulting school level ICC = 0.1), an ICC of 0.05 was chosen for two reasons: 1) the cohort was being randomised at the individual level within clusters (initially believed to be schools), so clustering was expected to have less of an effect within the analysis; and 2) restricting the cohort to the lowest achieving third of students nationally we believed would likely produce less variation between clusters, thus reducing the ICC. A correction factor of 1.45 resulted in an adjusted student sample of 131 students per cohort (Primary and Secondary).

3. Attrition

To account for the loss of students across the trial period (e.g., moving schools, dropout etc), an arbitrary value of 5% was added to the adjusted sample. The resulting sample was 137 students to be recruited from each cohort (Primary and Secondary).

Through the recruitment process, 135 students (67 intervention / 68 control) from 30 classes at 12 schools were recruited for the Primary cohort, and 169 students (85 intervention / 84 control) from 40 classes at 11 schools for the Secondary cohort. Table 1 presents the Minimum Detectable Effect Size (MDES) at different stages of the study for the combined cohort and Primary and Secondary subgroups. As students were randomised within multiple class groupings at each school (rather than within one class at the school), the class a student belonged to was used to examine the effect of

clustering within the sample. Linear mixed models used for outcome analysis also used the class a student belonged to for adjustment of clustering within models (see below).

A variance components model with the student (level 1) and the class (level 2) was used to establish the proportion of variance within the outcome variable (PAT-M) attributed to clustering (variance not explained at the student level that contributes to the ICC). This was considered a conservative approach as all variance not attributed to the student (class and school) was likely contained in the ICC when using the student's class, resulting in a higher ICC (and larger correction factor) than using the school as the cluster level. To maintain the most transparent research process, MDES was not modified using covariates (Bloom, Richburg-Hayes, & Black, 2007).

Table 1: Minimum detectable effect size at different stages

Stage	N [schools/ students] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES) *f, d*
Total							
Protocol - expected	24/274 (137; 137)	0.5	0.05	School, 24 schools	80%	0.05	0.20
Randomisation (PAT-M Baseline)	23/288 (146; 142)	0.5	0.34	Class, 70 classes	80%	0.05	0.24
Analysis (PAT-M time 3)	23/257 (130; 127)	0.6	0.18	Class, 68 classes	80%	0.05	0.18
Analysis (model ICC)	23/257 (130; 127)	0.6	0.09	Class, 68 classes	80%	0.05	0.16
Primary							
Protocol - expected	12/137 (72; 72)	0.5	0.05	School, 12 schools	80%	0.05	0.33
Randomisation (PAT-M Baseline)	12/133 (67; 66)	0.5	0.17	Class, 30 classes	80%	0.05	0.38
Analysis (PAT-M time 3)	12/119 (58; 61)	0.6	0.16	Class, 30 classes	80%	0.05	0.28
Analysis (Model ICC)	12/119 (58; 61)	0.6	0.10	Class, 30 classes	80%	0.05	0.26
Secondary							
Protocol	12/137 (72; 72)	0.5	0.05	School, 12 schools	80%	0.05	0.33
Randomisation (PAT-M Baseline)	11/155 (79; 76)	0.5	0.07	Class, 40 classes	80%	0.05	0.24
Analysis (PAT-M time 3)	11/138 (72; 66)	0.6	0.21	Class, 38 classes	80%	0.05	0.26
Analysis (Model ICC)	11/138 (72; 66)	0.6	0.06	Class, 38 classes	80%	0.05	0.24

Minimum detectable effect size was determined by back transformation of the recruited sample given the available information (pre-post correlation, ICC, cluster number and average cluster size) using the process described above. The back transformed raw sample (total sample / correction factor calculated using available information) was used in a sensitivity analysis in G*Power to determine the resulting effect size *f*. This effect size *f* was converted to effect size *d* for easier interpretation using the calculations in Cohen (1988).

Clustering at the class level produced higher ICC values than expected, however this was somewhat counteracted by a reduction in the average cluster size when calculating the correction factor. The MDES for the Primary and Secondary subgroups were considered slightly underpowered ($d > 0.2$), however when the cohorts were combined for the total analysis, the sample was considered sufficient to detect an effect as low as $d = 0.2$ when the ICC was produced using variance components of the final assessment value (PAT-M time 3) with no covariates used, or from the final hierarchical model containing covariates (Model ICC – see below for details).

2.5 Randomisation

Randomisation occurred after baseline assessment and took place in March 2017, using the non-scaled version of the primary outcome variable (PAT-M raw score). Randomisation was undertaken at the individual level within each class group, resulting in an intervention and a control group within each class group. Participants within each class were stratified by gender, and ranked within strata by their baseline PAT-M raw score. Gender based pairs were formed using their respective ranking (e.g., 1 and 2, & 3 and 4, etc). The highest ranked participant within a pair was randomised to one of two conditions first (intervention or 18-month wait-list control), with the other participant in the pair allocated to the alternate condition. A member of the University of Newcastle evaluation team (AM) carried out the randomisation procedure by tossing a coin. During randomisation, “heads” resulted in allocation to the intervention group, and “tails” the control group. The evaluator carrying out the randomisation procedure recorded the outcome against the name of each person within a class and the trial manager was responsible for communicating the results with the relevant school contact.

In the case of an uneven number of students within gender strata from an individual class (e.g., 3 girls and 5 boys), any participants not in a pair were randomised individually to one of the two conditions via a coin toss.

2.6 Analysis

Imbalance at baseline for analysed groups

Independent sample t-tests were used to evaluate whether random assignment resulted in equivalent groups at baseline for the primary and secondary outcomes. Gender proportions within groups were also calculated and the distribution of gender between groups assessed using Chi square analysis.

Primary outcome analysis

The primary aim of the analysis was to assess whether the QuickSmart intervention had a significant impact on students' mathematics achievement, as measured by the 18-month post-intervention PAT-M test scores. A linear model predicting 18-month post intervention PAT-M scores was fitted. Baseline PAT-M scores were included as fixed effects to control for prior achievement. Gender and stage (Year 4 or Year 8) were included as fixed effects to adjust for these covariates, and group (intervention or control) was included as a fixed effect to examine if group allocation had a significant effect on PAT-M results. The regular mathematics class a student belonged to was included as a random intercept within the model to account for clustering of students within classes. Statistical analyses were completed using PASW Statistics 21 (SPSS Inc. Chicago, IL) software. Impacts were estimated using an intention-to-treat protocol. Alpha levels were set at $p < 0.05$. Group means and 95% confidence intervals (CIs) were determined using the linear model specified below.

For the ith student in the jth class, let Y_{ij} be the student's PAT outcome at 18 months. The model equation is

$$Y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + r_{ij}$$

where $r_{ij} \sim N(0, \sigma^2)$ is random error and the intercept $\beta_{0j} = \gamma_{00} + u_{0j}$ consists of a random effect $u_{0j} \sim N(0, \tau_{00})$ of random variations around the overall mean γ_{00}. Binary predictors in the model are:

x_1 is Group allocation (Intervention or Control)

x_2 is Gender (Male or Female)

x_3 is Stage (Year 4 or Year 8)

and finally the model is adjusted for the student's baseline PAT score via the continuous covariate x_4.

Note. This analysis was modified from that originally specified in the statistical analysis plan. The model design was modified from a linear mixed model, to an ANCOVA style multi-level linear model in order to align with the common analysis practice of EEF and SVA.

Interim analysis

Analysis of the effects of the QuickSmart intervention without a delayed post-test (Term 2, 2018) were undertaken using the immediate follow-up PAT-M results (Term 4, 2017). The same linear model specified for the primary outcome analysis above, with the exception of let Y_{ij} as the student's PAT outcome at immediate follow-up.

Treatment effects in the presence of non-compliance
Instrumental variable analysis

A Two Stage Least Square (2SLS) approach was undertaken using the SYSLIN procedure in SAS. The first stage of the 2SLS involved regression of group allocation (intervention = 1; control = 0) on the continuous compliance instrument (proportion of QuickSmart sessions undertaken), with covariates: baseline PAT-M scaled score, gender and stage number. The second stage involved regression of the dependent variable (Follow-up PAT-M scaled score) on the predicted values obtained from the first stage, with covariates baseline PAT-M scaled score, gender and stage number.

This analysis was not pre-specified in the trial protocol.

Secondary outcome analysis

Secondary outcomes were assessed using an intention-to-treat protocol. The model detailed in the primary analysis was applied to the secondary outcomes.

Sub-group analysis

Primary and Secondary cohorts were investigated as sub-groups for analysis of primary outcomes. In each case, the statistical procedure for the analysis of the whole group analysis were followed, with the exception of the removal of the stage (Year 4 or Year 8) covariate from models.

Effect size calculation

Hedges' g was used to determine effect sizes of the change in mean score for each group relative to the baseline value (effect of intervention on the change score).

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s^*}$$

where our conditional estimate of $\bar{x}_1 - \bar{x}_2$ is recovered from $\beta 1$ in the primary ITT analysis model;

s^* is estimated from the analysis sample as follows:

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where $n1$ is the sample size in the control group, $n2$ is the sample size in the treatment group, $s1$ is the standard deviation of the control group, and $s2$ is the standard deviation of the treatment group (all estimates of standard deviation used are unconditional).

Ninety-five per cent confidence intervals (95% CIs) of the effect size were computed using the compute.es function in R. This function computes the confidence intervals using the variance in g derived by the Hedges & Olkin (Abel, Hannon, Mullineaux, & Beighle, 2011; 1985, p. 86) formula:

$$var(g) = \frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2)}$$

2.7 Implementation and fidelity

QuickSmart implementation and program fidelity were assessed using the following methods:

Student attendance (QuickSmart intervention exposure)

Program instructors for each of the QuickSmart Numeracy sessions were asked to maintain a record of student exposure (proportion of sessions attended).

Fidelity observations of QuickSmart sessions

In order to check intervention fidelity and identify any between-school differences in the delivery of the QuickSmart program, members of the evaluation team conducted observations of at least one QuickSmart lesson in 20 of the 23 evaluation schools (of which 7 were conducted jointly by two researchers). Sessions were evaluated against a 13-item checklist (Appendix C) developed in conjunction with the program developers in order to ensure that the program was delivered as intended. These observations were also used to establish a qualitative understanding of how students had experienced the program, as well as to contextualise the interview data. This process provided measures of consistency in program delivery and levels of student engagement across instructors.

Mathematics Teaching Efficacy Beliefs scale

To evaluate program instructors' confidence in delivering mathematics instruction to students, two scales from the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) were administered to instructors via an online survey post- intervention, during Term 4, 2017. Mathematics teacher efficacy (13 Questions), and mathematics teaching outcome expectancy (8 Questions) were used. This instrument is widely used with in-service and pre-service teachers, and has proven validity (Enochs, Smith & Huinker, 2000). Differences between efficacy measures among Primary and Secondary instructors were assessed using independent sample t-tests.

2.8 Process evaluation

Process evaluation was conducted throughout the intervention period in order to evaluate the feasibility of the QuickSmart program within each school setting (Primary and Secondary). Feasibility was assessed using the following methods:

Interviews

Student interviews

Interviews with a random sample of students were used to examine their experiences of the QuickSmart program. Two schools were selected per cohort via random number generation (1st two school numbers appearing in random generation list). Two intervention students were randomly selected from each school (as described above using subject numbers), along with two students in their classes randomised to the control group (*n* = 16 students). The selected students were invited to participate in pre-intervention and post-intervention interviews to explore their opinions about mathematics and their achievement in mathematics. These interviews were conducted by Dr Jess Harris and Wendy Taggart (UON).

Additional interviews were undertaken with the intervention condition students in the above group (*n* = 8). These interviews took place at three time-points (Weeks 8, 16, and 24) during the QuickSmart program to ascertain the experience of those participating in the QuickSmart program. Interviews with students participating in the program explored a range of topics, including:

- student experiences of learning mathematics;
- student perceptions of being asked to participate in the program;
- student perceptions of participating in the program;
- student reports of the social impact of being withdrawn from class;
- student experiences of the program (what happens in the program; does the process/ interaction change depending on who is running the program?); and,
- whether student experience of the program changes over time.

Post-intervention interviews were conducted with students from both the control and intervention conditions from the four randomly selected schools. One Secondary student from the intervention condition declined a post-intervention interview. Two students in the intervention condition were absent on the day of the interviews at one time point each and so the data for these students were limited to interviews at three time points. Two other students, who had undertaken QuickSmart in the school, along with the students in the control condition, who had been interviewed pre-intervention, agreed to a post-intervention interview to share their experiences.

Teacher interviews

Post-intervention interviews were conducted with a randomly selected sub-sample (random number generation using class number) of six classroom teachers (four Primary and two Secondary). These interviews gathered teachers' perceptions of:

- outcomes of the program for their students;
- the social impact on students of being involved in the program; and
- whether their students' experiences of the program changed over time.

These interviews were conducted by Dr Jess Harris (UON).

QuickSmart instructor interviews

Interviews were conducted with QuickSmart instructors (n = 5) who delivered the program to the students involved in the interview phase. These interviews aimed to examine key barriers and enablers to their effective delivery of the QuickSmart program. These interviews were conducted by Dr Jess Harris (UON).

Data analysis

Qualitative analysis of the student, instructor and teacher interviews was conducted, with all interview transcriptions read by at least two members of the evaluation team. Transcripts were coded deductively and inductively to identify themes relating to the concepts detailed above in Outcome measures.

3 Impact evaluation

3.1 Timeline

The intervention commenced at slightly different times at each of the schools involved due to student recruitment and assessment only able to start in Term 1, 2017 as new classes were formed and finalised at each school (Table 2). All schools were assessed and participants randomised prior to the end of Term 1, 2017. This timing gave schools the opportunity to achieve the full 30 week schedule required for the QuickSmart program. Some schools, however, opted to delay starting beyond randomisation and commence the QuickSmart program at the start of Term 2, 2017.

Table 2: Study timeline

Time period	Activity
Terms 3-4, 2016 – Term 1, 2017	School recruitment (SCS)
Term 1, 2017	Baseline data collection (UON) • Primary outcome measure (PAT-M) • Student mathematics self-belief measures (PISA) • Student interviews (both conditions) • Instructor/teacher efficacy measures (MTEBI) Training workshop 1 (SiMERR) Schools commence QuickSmart intervention QuickSmart program instructor adherence (self-report checklists)
Term 2, 2017	Schools continue QuickSmart intervention QuickSmart program instructor adherence (self-report checklists) Observations of QuickSmart sessions (UON) Student interviews (intervention condition – UON)
Term 3, 2017	Schools continue QuickSmart intervention QuickSmart program instructor adherence (self-report checklists) Observations of QuickSmart sessions (UON) Student interviews (intervention condition – UON) Training workshop 2 (SiMERR)
Term 4, 2017	Schools continue QuickSmart intervention QuickSmart program instructor adherence (self-report checklists) Observations of QuickSmart sessions (UNE) Student interviews (intervention condition – UNE) Training workshop 3 (SiMERR) Post-intervention data collection (UON): • Primary outcome measure (PAT-M) • Student self-belief measures (PISA) • Student interviews (both conditions) • Instructor/teacher efficacy measures (MTEBI)
Term 2, 2018	Follow-up data collection (UON) • Primary outcome measure (PAT-M) • Student self-belief measures (PISA) Evaluation data analysis Wait-list control schools commence QuickSmart intervention
Term 3, 2018	Evaluation data analysis and final reporting

3.2 School and Instructor characteristics

The Index of Community Socio-Educational Advantage (ICSEA) provides an indication of the socio-educational backgrounds of students in each school. The majority of schools involved in the evaluation displayed ICSEA values marginally above the mean national value of 1000 (Table 3) (Australian Curriculum Assessment and Reporting Authority, 2011).

Table 3: School ICSEA by cohort

Cohort	Below 1000 (N, %)	Above 1000 (N, %)	ICSEA Mean (95% CI)
Primary	1, 8%	11, 91%	1054 (1020, 1088)
Secondary	3, 27%	8, 73%	1028 (1006, 1051)

The majority of schools (65.2%) recruited to the trial were established QuickSmart schools, having previously undertaking or currently delivering QuickSmart (Table 4). Only one new school was recruited within the Primary cohort, with the majority of new schools recruited within the Secondary cohort, making up the majority of schools within this cohort (63.6%) undertaking QuickSmart for the first time.

Table 4: School QuickSmart status by cohort

QuickSmart status	Primary	QuickSmart status	Primary
Existing	11 (91.7%)	4 (36.4%)	15 (65.2%)
New	1 (8.3%)	7 (63.6%)	8 (34.8%)
Total	**12**	**11**	**23**

Individual schools recruit and employ QuickSmart instructors to deliver the program. This model was followed during the evaluation to replicate the regular experience of schools. The majority of QuickSmart instructors for both cohorts were qualified TA, with a small proportion from each cohort holding graduate or post-graduate teaching qualifications, and several holding no teaching qualification (Table 5). Instructors in the Primary cohort were highly experienced, with the majority (>70%) reporting more than three years of instructing experience (Table 6). The majority of Primary instructors (57%) had also completed the most advanced level of training (Table 7). In comparison, the Secondary instructors were in their first year of instruction (68%), and had only completed the first level of training (74%).

Table 5: Instructor qualifications

Qualification	Primary N (%)	Secondary N (%)	Total N (%)
Teaching assistant certificate	13 (61.9%)	8 (42.1%)	21 (52.5%)
Teaching degree	2 (9.5%)	3 (15.8%)	5 (12.5%)
No qualification	2 (9.5%)	3 (15.8%)	5 (12.5%)
Post-graduate teaching	1 (4.8%)	2 (10.5%)	3 (7.5%)
Missing	3 (14.3%)	3 (15.8%)	6 (15.0%)
Total	**21**	**19**	**40**

Table 6: Instructor experience

Experience	Primary N (%)	Secondary N (%)	Total N (%)
1st year	3 (14.3%)	13 (68.4%)	16 (40.0%)
1-2 years	0 (0.0%)	1 (5.3%)	1 (2.5%)
3-4 years	6 (28.6%)	1 (5.3%)	7 (17.5%)
5 or more years	9 (42.9%)	1 (5.3%)	10 (25.0%)
Missing	3 (14.3%)	3 (15.8%)	6 (15.0%)
Total	**21**	**19**	**40**

Table 7: Instructor training

Training	Primary N (%)	Secondary N (%)	Total N (%)
None	0 (0.0%)	1 (5.3%)	1 (2.5%)
Basic Training (Year 1)	5 (23.8%)	14 (73.7%)	19 (47.5%)
Advanced (Year 2)	1 (4.8%	1 (5.3%)	2 (5.0%)
Advanced (Year 3)	12 (57.1%)	0 (0.0%)	12 (30.0%)
Missing	3 (14.3%)	3 (15.8%)	6 (15.0%)
Total	**21**	**19**	**40**

Mathematics Teaching Efficacy Beliefs scale

Instructors for the Primary cohort displayed significantly greater (t = 2.49, df = 28, p = 0.02) efficacy towards teaching mathematics than the Secondary instructors (mean difference; 95%CI: 0.39; 0.07 – 0.72) on average (Table 8). Primary instructors also displayed higher, but not statistically significant values (t = 1.16, df = 31, p = 0.25) for outcome expectancy than the Secondary instructors. This result likely reflects the differences in experience and training between the two cohorts.

Table 8: Mathematics efficacy beliefs scale* by cohort – end of intervention period

	Primary		Secondary	
Outcome	**N**	**Mean (95% CI)**	**N**	**Mean (95% CI)**
Mathematics Teaching Efficacy scale	18	4.28 (4.06, 4.49)	15	3.88 (3.62, 4.15)
Outcome Expectancy scale	18	3.52 (3.21, 3.82)	15	3.27 (2.96, 3.60)

*Scale range 1 – 5

3.3 Student characteristics

There was a higher proportion of female participants in the evaluation, with no significant difference ($p > 0.05$) between the gender proportion between randomised groups (Table 9).

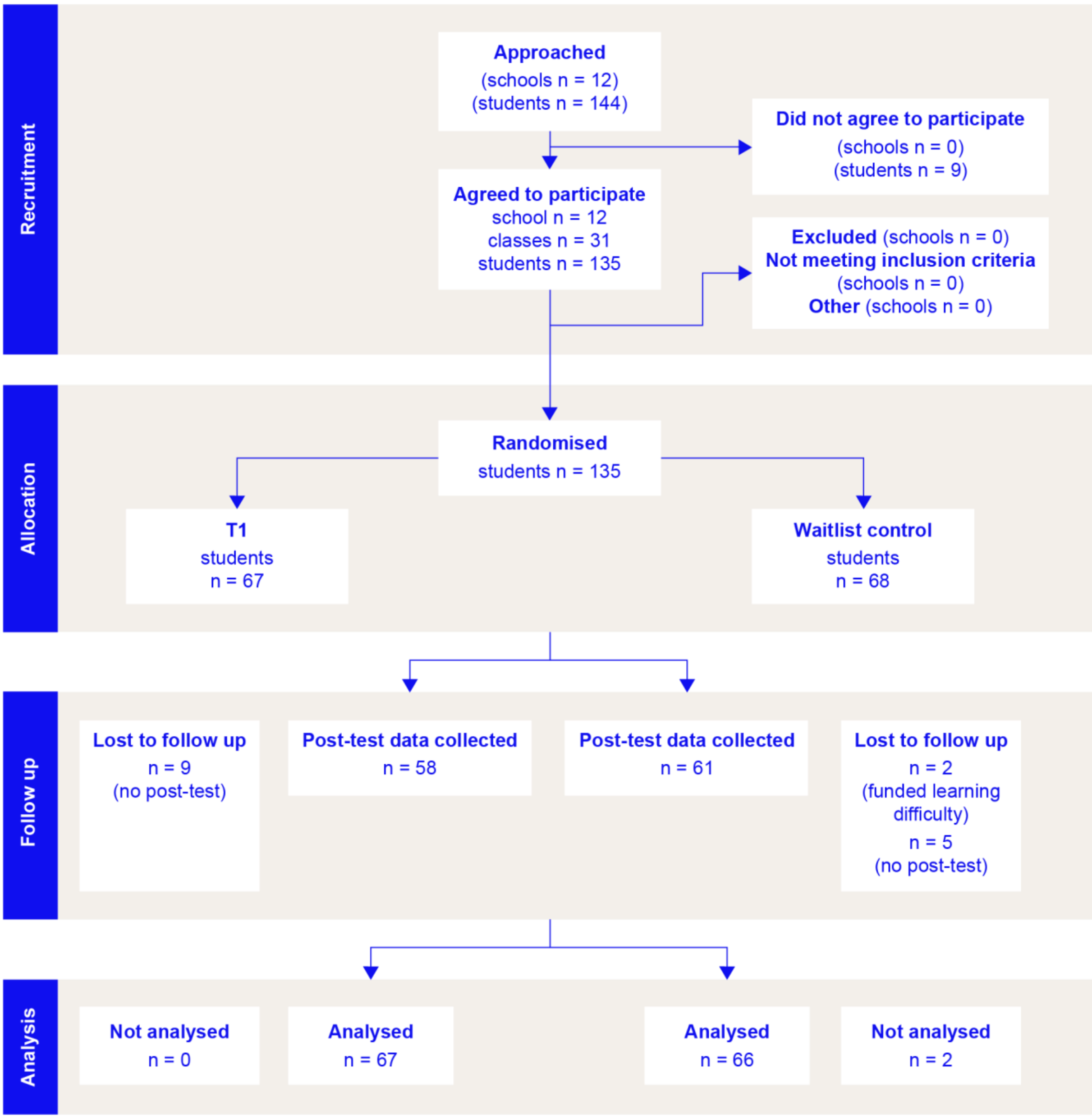
Table 9: Gender by cohort and randomised group

Characteristic	Intervention		Control		Bias	p
Student (categorical)	**n/N (missing)**	**%**	**n/N (missing)**	**%**		**Chi-square**
Total (female)	87/146	59.6	88/142	62.0	-2.4	0.68
Primary (female)	37/67	55.2	39/66	59.1	-3.9	0.65
Secondary (female)	50/79	63.3	49/76	64.5	-1.2	0.88

Participant flow including losses and exclusions

Participant flow is detailed in Figure 2. All of the suitable schools identified and approached by SCS consented to participate, with 94% and 97% of the invited students from the Primary and Secondary cohorts consenting to involvement respectively. Two control group students from the Primary cohort (<2%) and 14 students (6 intervention, 8 control) from the Secondary cohort (8%) were diagnosed with a learning disability during the trial period. These students continued with the QuickSmart intervention, but were excluded from all analyses as funded learning difficulties were listed as exclusion criteria for invitation into the study. These students have been removed from the participant flow diagram (Figure 2).

Figure 2: Participant flow diagram – Total cohort



Missing data

Post test data were not available for 14 students (9 intervention, 5 control) in the Primary cohort (10.5%), and 16 students (6 intervention, 10 control) in the Secondary cohort (10.4%). These students either withdrew from the study (Primary n = 0, Secondary n = 4), moved to a different school (Primary n = 13, Secondary n = 8), or were absent on the day of testing (Primary n = 1, Secondary n = 4) during the study. Overall, 10.4% (n = 30/ 288) of cases for the PAT-M scaled score were lost to follow-up. The logistic regression procedure described in the SAP for determining the mechanism of missingness was not significant, indicating the mechanism was not MAR. Given that 21/ 30 (7.1%) and 5/ 30 (1.7%) missing cases at follow-up were due to students moving school and sickness on the day of testing respectively, with only 4/ 30 (1.4%) missing cases due to study withdrawal, the missing data was assumed to be MCAR. Imputation was not undertaken for any missing data.

One participant (Secondary cohort) was excluded from the primary analysis as an extreme outlier (>2.5 standard deviations below the mean).

An independent sample t-test evaluating the baseline primary outcome (PAT-M) displayed no significant difference (t = 0.71, df = 286, p = 0.48) between those lost-to-follow up and the remaining recruited participants (Mean difference; 95%CI = 0.72; -1.28 – 2.74)

3.4 Outcomes and analysis

Baseline characteristics

At baseline, there was no significant difference identified between groups (intervention vs control) for the primary outcome for the total cohort, or among Primary and Secondary sub-groups (Table 10). Among secondary outcome variables, there was no significant differences identified between groups for the total cohort on any of the outcome variables (Table 11). Primary and Secondary sub-groups also displayed the same result for Secondary outcome variables (Table 12 and Table 13).

Table 10: Baseline comparison of primary outcome (PAT-M)

	Intervention group		Control group		Effect size	
Outcome	**N**	**Mean (95% CI)**	**N**	**Mean (95% CI)**	**Hedges g (95% CI)**	**p-value**
Total	146	45.40 (44.20, 46.59)	142	45.52 (44.30, 46.73)	-0.02 (-0.25, 0.22)	0.89
Primary	67	41.25 (39.60, 42.90)	66	41.74 (40.08, 43.41)	-0.07 (-0.41, 0.27)	0.68
Secondary	79	48.92 (47.60, 50.23)	76	48.79 (47.45, 50.14)	-0.01 (-0.33, 0.31)	0.90

Table 11: Baseline comparison of secondary outcomes – Total cohort

	Intervention group		Control group		Effect size	
Outcome	N	Mean (95% CI)	N	Mean (95% CI)	Hedges g (95% CI)	p-value
Maths self-efficacy	146	2.65 (2.55, 2.74)	142	2.64 (2.55, 2.74)	0.02 (-0.21, 0.25)	0.96
Maths self-concept	146	2.35 (2.24, 2.47)	141	2.39 (2.27, 2.50)	-0.06 (-0.29, 0.17)	0.71
Maths interest	145	2.58 (2.45, 2.72)	142	2.64 (2.50, 2.78)	-0.07 (-0.30, 0.16)	0.57
Maths anxiety	146	2.39 (2.26, 2.53)	141	2.45 (2.32, 2.59)	-0.07 (-0.31, 0.16)	0.54

Table 12: Baseline comparison of secondary outcomes – Primary cohort

	Intervention group		Control group		Effect size	
Outcome	**N**	**Mean (95% CI)**	**N**	**Mean (95% CI)**	**Hedges g (95% CI)**	**p-value**
Maths self-efficacy	67	2.92 (2.79, 3.04)	66	2.89 (2.75, 3.02)	0.06 (-0.28, 0.40)	0.74
Maths self-concept	67	2.53 (2.41, 2.66)	66	2.57 (2.45, 2.70)	-0.07 (-0.41, 0.28)	0.61
Maths interest	67	2.90 (2.72, 3.08)	66	3.03 (2.86, 3.21)	-0.19 (-0.53, 0.16)	0.30
Maths anxiety	67	2.52 (2.31, 2.72)	66	2.58 (2.37, 2.78)	-0.07 (-0.42, 0.27)	0.68

Table 13: Baseline comparison of secondary outcomes – Secondary cohort

	Intervention group		Control group		Effect size	
Outcome	**N**	**Mean (95% CI)**	**N**	**Mean (95% CI)**	**Hedges g (95% CI)**	**P**
Maths self-efficacy	79	2.42 (2.30, 2.54)	76	2.43 (2.31, 2.55)	-0.02 (-0.34, 0.30)	0.96
Maths self-concept	79	2.18 (2.07, 2.29)	76	2.16 (2.04, 2.27)	0.03 (-0.29, 0.35)	0.76
Maths interest	79	2.31 (2.13, 2.50)	76	2.30 (2.11, 2.48)	0.01 (-0.31, 0.33)	0.96
Maths anxiety	79	2.30 (2.13, 2.48)	76	2.33 (2.15, 2.51)	-0.04 (-0.36, 0.28)	0.81

Program exposure

The average (SD) number of the desired 90 sessions undertaken for those with data available at immediate post-test was 60 (SD = 19) sessions, equating to an average exposure of 54.9% of the desired 90 sessions. Exposure among Primary and Secondary cohorts differed greatly, with average sessions attended at 66 (SD = 12) and 46 (SD = 17) for the Primary and Secondary cohorts respectively. This equated to 73% and 49% of the recommended intervention volume in the intervention period. For the purposes of this effectiveness trial, these average exposures were in line with the mean of average yearly exposure of 66%, and 55% for Year-4 and Year-8 cohorts respectively reported by SiMERR from 2013 – 2016 (Pegg et al., 2013, 2014, 2015, 2016).

Due to recruitment and testing processes, not all schools received access to the full 30-week intervention period required to undertake 90 sessions (three sessions per week). School level average exposure (%), the intervention time available to the schools (weeks), and the QuickSmart status of the school (new or existing school) is outlined in Tables 14 and 15 for Primary and Secondary cohorts respectively, with schools rank in descending order by their average QuickSmart exposure (%). Average exposure varied widely across schools within each cohort, with no clear trend evident regarding the weeks available and average exposure.

When considering if a school was a new or existing QuickSmart school in this trial (Table 16), the one new school in the Primary cohort displayed 10% lower average exposure than established schools,

however this only had an effect on three intervention participants at this school. The Secondary cohort displayed very little difference in average exposure between new and existing schools in this trial, with 0.4% difference between new and existing schools.

Table 14: Average intervention exposure per school - Primary

School	QS exposure % Mean (SD)	Available intervention weeks	N	QS status
7	94.8 (8.4)	30	6	existing
14	85.6 (12.2)	28	6	existing
9	84.3 (3.9)	30	6	existing
3	76.9 (5.1)	30	6	existing
21	74.4 (4.3)	29	6	existing
1	69.8 (3.4)	30	5	existing
4	68.4 (4.9)	30	5	existing
6	68.2 (7.1)	30	5	existing
8	65.8 (2.9)	30	5	existing
20	62.6 (2.8)	23	3	new
11	62.1 (6.1)	30	6	existing
2	53.1 (3.5)	28	6	existing
Total	**72.8 (12.9)**		**65**	

Table 15: Average intervention exposure per school - Secondary

School	QS exposure % Mean (SD)	Available intervention weeks	N	QS status
16	72.9 (5.4)	29	8	existing
18	63.2 (8.1)	30	8	new
22	59.1 (1.9)	28	6	new
13	56.7 (5.1)	30	8	new
23	52.6 (6.3)	29	6	existing
17	52.2 (12.4)	28	5	new
15	44.2 (15.4)	27	7	new
19	42.6 (7.2)	30	8	new
10	41.1 (6.3)	30	2	existing
12	26.7 (6.5)	28	8	existing
5	26.2 (13.6)	26	9	new
Total	**49.4 (16.7)**		**75**	

Table 16: Average intervention exposure by QuickSmart status

QS status	New		Existing		
	N	QS exposure % Mean (SD)	N	QS exposure % Mean (SD)	Difference
Primary	3	62.6 (2.8)	62	73.3 (13.1)	10.7
Secondary	51	49.3 (14.8)	24	49.7 (20.1)	0.4

Intervention exposure is reported in quartiles in Table 17 for the Primary cohort, and Table 18 for the Secondary cohort. In the Primary cohort, 35% of students completed between 75%-100% of the full program, and in the Secondary cohort, 4% of students reached this level.

Table 17: Intervention exposure – Primary cohort

Exposure	Frequency	Percent (%)
0 - 24%	0	0%
25 - 49%	2	3.1%
50 - 74%	40	61.5%
75 - 100%	23	35.4%
Total	**65**	**100.0%**

Table 18: Intervention exposure – Secondary cohort

Exposure	Frequency	Percent (%)
0 - 24%	7	9.3%
25 - 49%	27	36.0%
50 - 74%	38	50.7%
75 - 100%	3	4.0%
Total	**75**	**100.0%**

Primary outcome analysis

The purpose of the primary analysis of this evaluation was to determine if there was significant difference in the progress made by students in the QuickSmart program when compared against a control group receiving only regular classroom mathematics instruction. There was no significant positive effect (g = 0.05, $p = 0.59$) displayed for the PAT-M results of the QuickSmart intervention group in comparison to the control group (Table 19). Analysis of mathematics achievement (PAT-M) by cohort identified no significant effect on PAT-M results within the Primary ($g = 0.08$, $p = 0.48$) or Secondary ($g = 0.01$, $p = 0.95$) cohorts.

Table 19: Primary outcome analysis (PAT-M) – 18-month follow-up

Cohort	Intervention group		Control group		Effect size		
	n (miss)	Mean (95% CI)	n (miss)	Mean (95% CI)	*B* (95% CI)	Hedges g (95% CI)	P
Primary analysis							
PAT-M (18 months)	130 (15)	52.44 (51.32, 53.55)	127 (15)	52.06 (50.94, 53.19)	0.38 (-0.99, 1.74)	0.05 (-0.19, 0.30)	0.59
Subgroup analysis							
Primary	58 (9)	52.79 (50.91, 54.68)	61 (5)	52.02 (50.17, 53.87)	0.77 (-1.38, 2.93)	0.08 (-0.28, 0.44)	0.48
Secondary	72 (6)	52.13 (50.75, 53.49)	66 (10)	52.07 (50.65, 53.48)	0.06 (-1.70, 1.83)	0.01 (-0.33, 0.35)	0.95

Interim analysis

Analysis of the effects of the QuickSmart intervention without a delayed post-test (Term 2, 2018) were undertaken using the immediate follow-up PAT-M results (Term 4, 2017) (Table 20). Whilst the effect sizes were larger at this time point, there was no significant positive effect on PAT-M results (g = 0.09, p = 0.30) displayed among the whole cohort, or for either the Primary (g = 0.10, p = 0.42) or Secondary (g = 0.08, p = 0.52) subgroups.

Table 20: Primary outcome analysis (PAT-M) – Intervention post test

Cohort	Intervention group		Control group		Effect size		
	n (miss)	Mean (95% CI)	n (miss)	Mean (95% CI)	*B* (95% CI)	Hedges g (95% CI)	P
Primary analysis							
PAT-M (Post-test)	139 (6)	50.61 (49.57, 51.65)	137 (5)	49.86 (48.82, 50.91)	0.74 (-0.67, 2.17)	0.09 (-0.15, 0.33)	0.30
Subgroup analysis							
Primary	65 (2)	49.85 (48.08, 51.61)	65 (1)	48.85 (47.08, 50.62)	1.0 (-1.44, 3.43)	0.10 (-0.24, 0.45)	0.42
Secondary	74 (4)	51.29 (50.09, 52.49)	72 (4)	50.75 (49.54, 51.97)	0.53 (-1.09, 2.15)	0.08 (-0.25, 0.41)	0.52

Treatment effects in the presence of non-compliance

Instrumental Variable analysis

This analysis removes the contamination in the treatment effect of not getting the full intervention dosage by adjusting the parameter estimate for intervention compliance. Group allocation is adjusted for the level and intervention obtained (Stage 1), with the adjusted group allocation values entered in the place of group in the regression against the PAT-M outcome variable (Stage 2). This process results in a compliance adjusted estimate of effect.

Results from the first stage are reported in Table 21, with the correlation between group allocation and compliance variables reported alongside the item t, and model F test. The suggested rule of thumb for checking the strength of the instrument is that the F-statistic from the first stage regression should be greater than 10, or the t value of the instrument above approximately 3 (Angrist, 2006), otherwise the instrument is considered to be weak, the consequence of which is that the sampling distribution of the 2SLS estimator might not be approximately normal even in large samples. Values from the continuous QuickSmart compliance instrument (% of program) were considered to fulfil the criteria for a strong instrument.

The parameter estimates and corresponding effect size from the second stage are reported in Table 22. The compliance adjusted parameter estimate was 1.11 (95% CI: -0.47 – 2.69), increasing by 0.73 from that of the primary analysis (*B* = 0.38; 95% CI: -0.99 – 1.74). Intervention effects adjusted for compliance were not found to be statistically significant, however there was a small increase in the effect of the QuickSmart intervention, equivalent to around one month's growth for the intervention above that of the control group.

Note. This analysis was not pre-specified in the original trial protocol.

Table 21: Stage one results – Model, item and instrument correlation (with group allocation variable)

Instrument	N	Model					Item		Correlation	
		DF	Sum of Squares	Mean Square	F Value	*P*	*t*	*P*	Pearson	*P*
Exposure %	257	4	54.36	13.59	354.43	<0.01	36.62	<0.01	0.87	<0.01

Table 22: Stage two results – Parameter estimates and effect sizes for the tested compliance instrument

Instrument	N	Estimate. *B* (95% CI)	Std. Error	t Value	*P*	Hedges g (95% CI)
Exposure %	257	1.11 (-0.47, 2.69)	0.81	1.37	0.17	0.09 (-0.16, 0.33)

Secondary outcome analysis

Student cognitive and affective measures

There were no significant intervention effects for any of the secondary outcomes (Table 23). Among the Primary cohort, there was no significant change in students' mathematics self-efficacy (g = 0.09, p = 0.65), however there was a significant positive effect on students' mathematics self-concept (g = 0.30, p = 0.04), and maths interest (g = 0.47, p = 0.01) (Table 24). This finding was supported by interview data (described in the process evaluation section, below), in which classroom teachers reported that students who had participated in the QuickSmart program appeared to gain the confidence to contribute more in their mathematics classroom. The mathematics anxiety measure was 0.23 scale units higher on average among intervention students in the Primary cohort, however this difference was not significant (g = 0.27, p = 0.06). There were no significant intervention effects for any of the secondary outcomes for the Secondary cohort (Table 25).

Table 23: Secondary outcomes

Cohort	Intervention group		Control group		Effect size		
	n (miss)	Mean (95% CI)	n (miss)	Mean (95% CI)	*B* (95% CI)	Hedges g (95% CI)	P
Maths self-efficacy	130 (15)	2.86 (2.77, 2.95)	126 (16)	2.83 (2.74, 2.92)	0.03 (-0.08, 0.14)	0.05 (-0.20, 0.30)	0.62
Maths self-concept	131 (14)	2.39 (2.29, 2.49)	123 (19)	2.29 (2.20, 2.40)	0.09 (-0.04, 0.22)	0.15 (-0.10, 0.39)	0.17
Maths interest	130 (15)	2.56 (2.44, 2.67)	125 (17)	2.45 (2.33, 2.57)	0.11 (-0.04, 0.26)	0.14 (-0.11, 0.38)	0.16
Maths anxiety	124 (21)	2.57 (2.40, 2.69)	131 (11)	2.51 (2.38, 2.63)	0.06 (-0.11, 0.22)	0.07 (-0.17, 0.32)	0.48

Table 24: Secondary outcomes – Primary cohort

Cohort	Intervention group		Control group		Effect size		
	n (miss)	Mean (95% CI)	n (miss)	Mean (95% CI)	*B* (95% CI)	Hedges g (95% CI)	P
Maths self-efficacy	58 (9)	3.14 (3.00, 3.27)	61 (5)	3.09 (2.96, 3.23)	0.04 (-0.13, 0.21)	0.09 (-0.26, 0.44)	0.65
Maths self-concept	58 (9)	2.67 (2.50, 2.84)	59 (7)	2.46 (2.29, 2.62)	0.21 (0.13, 0.41)	0.30 (-0.06, 0.65)	0.04
Maths interest	58 (9)	2.94 (2.76, 3.13)	59 (7)	2.57 (2.39, 2.75)	0.37 (0.15, 0.59)	0.47 (0.12, 0.83)	0.01
Maths anxiety	58 (9)	2.78 (2.56, 3.00)	60 (6)	2.55 (2.33, 2.76)	0.23 (-0.01, 0.47)	0.27 (-0.08, 0.62)	0.06

Table 25: Secondary outcomes – Secondary cohort

Cohort	Intervention group		Control group		Effect size		
	n (miss)	Mean (95% CI)	n (miss)	Mean (95% CI)	*B* (95% CI)	Hedges g (95% CI)	P
Maths self-efficacy	72 (6)	2.59 (2.47, 2.72)	65 (11)	2.58 (2.46, 2.71)	0.01 (-0.14, 0.16)	0.02 (-0.32, 0.35)	0.89
Maths self-concept	72 (6)	2.14 (2.03, 2.26)	64 (12)	2.16 (2.05, 2.28)	-0.02 (0.13, 0.41)	-0.03 (-0.37, 0.31)	0.81
Maths interest	72 (6)	2.23 (2.08, 2.38)	65 (11)	2.36 (2.21, 2.51)	-0.13 (-0.32, 0.07)	-0.18 (-0.51, 0.16)	0.19
Maths anxiety	72 (6)	2.38 (2.24, 2.52)	64 (12)	2.49 (2.34, 2.64)	-0.11 (-0.31, 0.09)	-0.15 (-0.49, 0.19)	0.29

3.5 Implementation and fidelity

Student attendance (QuickSmart intervention exposure)

As outlined in Tables 17 and 18, student exposure to the program varied. In the Primary cohort, 34% of students completed between 75%-100% of the full program, and in the Secondary cohort, 4% of students reached this level.

Fidelity of QuickSmart implementation

As a measure of assessing adherence to the program, observations of QuickSmart sessions were conducted at 20 of the 23 schools by UON researchers throughout the intervention period. Generally, the instructors were observed to be supportive and encouraging. Students mostly appeared engaged, comfortable and on-task. Many schools had a private, dedicated area to use for sessions (some sharing with other support roles), and some had particularly vibrant and well-organised spaces established. However, fidelity observations identified marked differences in the practices used by QuickSmart Instructors.

While the majority of QuickSmart Instructors implemented the program as designed, their approaches to each of the activities and to support student learning within the activities varied widely. Of the 22 individual sessions observed, all 13 items on the fidelity checklist were only observed in three cases; however, for 20 of the observations, at least 10 of the 13 items were observed. The most common activities not included in the observed sessions were independent worksheets, independent worksheet feedback, and the game normally conducted at the end of each session.

The evaluation team observed a number of between-school differences in the delivery of the QuickSmart program, which are summarised below:

- use of 'non-QuickSmart' resources – for example, use of other commercial resources in the independent worksheet activity;
- homework provided by some instructors;
- inconsistency with some activities undertaken by students and/or instructors in sessions - for example, marking their own or each other's work, instructors or students administering flashcards;
- some instructors commented that they modified the program to suit their purposes - for example, one instructor commented that she started all students at the first level, not at the level

recommended by the pre-testing as she thinks the standard program starting levels are too difficult;

- inconsistency with level of support provided during OZCAAS testing - some instructors used standard 'normal test conditions'; others actively assisted students throughout the tests;
- some instructors discussed the addition of 'problem-solving activities' in one session per week throughout the intervention period, however the nature of this work seemed unclear; for example, one school used its own template for problem-solving which appeared to be a guide for working out NAPLAN-style 'word' problems, not specifically concerning numerical operations and strategies;
- differences in preparation techniques at the start of sessions; and,
- inconsistent use of testing equipment.

The research team also witnessed that on occasion some QuickSmart Instructors encountered difficulties in seeking alternative ways to promote conceptual understanding in students when memorisation alone was not enough to progress through the program.

3.6 Cost

The estimated cost per student to participate in the QuickSmart program was calculated based on information provided by the program developers, as well as a random sample of 4 of the 23 schools (two Primary and two Secondary schools). The estimate is based on the intervention continuing at a school over a three-year period to reflect the typical training and licensing arrangements for the QuickSmart program. This estimate includes:

- usual licensing and equipment costs for the QuickSmart program;
- typical printing and stationery costs.

Analysis of cost per student does NOT include:

- direct staffing costs of Learning Support Officer/s and program coordinator;
- teacher release for training;
- equipment required to set up a QuickSmart learning space, as typically this is sourced from within the school; and
- administration costs from the intervention team during this evaluation.

Cost of purchasing the QuickSmart program

The initial cost of the QuickSmart program is $10,500 exc. GST payable at start-up. A QuickSmart licence of $7,000 exc. GST gives access to six days of training in the first year (for up to 5 staff), access to online resources and telephone support. Optional training during the second and third years can be arranged at an additional cost (See below). An resources cost of $3,500 exc. GST covers the required equipment, resources and a three-year OZCAAS testing licence, also payable at start-up. If a school continues to run QuickSmart past three years, an additional three-year OZCAAS licence will cost $1,800 (there is no additional QuickSmart licencing costs).

Evaluation schools

The 23 schools involved in this evaluation were allocated funding of $8,600 to support the implementation of the QuickSmart program. Schools were not required to pay the standard start-up fees (and the eight schools new to QuickSmart were provided with the required equipment and resources at no cost).

Estimated cost per student – excluding staffing costs

The cost per student is estimated at $151 per year, based on 25 students per year undertaking the intervention, and the program continuing over a three-year period at a school. This estimate includes:

- typical licensing and equipment costs for the program ($140 per student); and
- printing and stationery costs ($11)

Notes:

1. The above cost estimate does not include staffing costs; for example, of Learning Support Officer/s who deliver the QuickSmart sessions, or a QuickSmart program coordinator.
2. The above cost estimate does not include equipment required to set up a QuickSmart learning space, because this is typically sourced from within the school.
3. Six days of teacher release for each Learning Support Officer and/or coordinator is required to attend QuickSmart training in year one of the program.
4. Schools have the option to send staff to additional training days at an additional cost (approximately $3000 and three days of teacher release per participant in year 2, and $1000 and one day of teacher release per participant in year 3).
5. The cost per student will vary depending on numbers of students participating in the QuickSmart program. For example:

No. students participating in QuickSmart program	Estimated cost per student per year (if QuickSmart program continues for three years)
10	$361
20	$186
30	$128

Estimated cost per student – including staffing costs

When staffing costs are included in the cost analysis, the cost per pupil is estimated at $1,007 per year, based on 25 students per year undertaking the intervention, and the program continuing over a three-year period at a school. This estimate includes:

- typical licensing and equipment costs for the program ($140 per student);
- direct staffing costs of a Learning Support Officer (24hrs/student @ $30 = $720 per student) and program coordinator (4hrs/student @ $30 = $120 per student);
- learning support officer release for training ($16); and
- printing and stationery costs ($11)

Notes:

1. The above cost estimate does not include equipment required to set up a QuickSmart learning space (computer and stationary), because this is typically sourced from within the school.
2. Schools have the option to send staff to additional training days at an additional cost ($3,000 per school for the three days of training, and three days of teacher release per instructor attending the training in year 2, and $1000 per school, and one day of teacher release per instructor attending in year 3) – this is not included in this costing.

4 Process evaluation

Key data used to inform the process evaluation were interviews with QuickSmart instructors, classroom teachers and students regarding the feasibility and success of the program. Interviews and fidelity observations of QuickSmart sessions provided valuable information that supported the quantitative data and highlighted several issues that could threaten the feasibility of program delivery within the school setting.

4.1 Perceptions of program success

Interest in and confidence with mathematics

Qualitative data drawn from interviews with students from both the control and intervention conditions identified that students selected for participation in the evaluation held overwhelmingly negative attitudes to mathematics and their prior achievement in the subjects. No change in students' attitudes to mathematics was identified during the evaluation process.

> *I definitely don't really like maths. It's not my strongest subject.*
> *(Year 4 student, pre-intervention interview)*

> *I'm saying I'm dumb because I really don't like maths and I just don't know anything in maths.*
> *(Year 4 student, pre-intervention interview)*

When considering the impact of QuickSmart on students, Primary teachers reported their belief that students who participated in the program showed greater confidence in the classroom. Substantial research confirms the existence of a relationship between confidence and achievement in mathematics (Ma, 1999). According to this research, students who demonstrate greater interest and willingness to try in mathematics classrooms are more likely to improve their performance in mathematics.

Interviewed Primary teachers reported noticing a shift in students' willingness to contribute in their maths classes, even when they are not sure whether their responses are correct.

> *I think this has absolutely helped with them offering answers because it's okay to make mistakes with QuickSmart too. You fix it. It's not the end of the world to make a mistake.*
> *(Year 4 Maths Teacher, post-intervention interview)*

The interviews corroborate survey findings with regards to mathematics self-concept and interest for the Primary intervention group, which found a statistically significant improvement in student self-belief on these two scales in comparison to the control group. In contrast, survey data from Secondary students in the intervention condition did not demonstrate any significant effects on the four measures of students' mathematics self-concept, self-efficacy, interest, and anxiety.

Small-group interactions

Teachers also suggested that the small group environment provided by QuickSmart holds benefits for student learning. Mathematics teachers, QuickSmart Instructors and students all reported that working with one other student, whose mathematics achievement was similar to their own, provided a less confronting learning environment than their regular classroom.

> *I just like it because like I'm actually learning like that.*
> *(Year 4 student, early-intervention interview)*

Students in the QuickSmart program reported that they were answering mathematics questions more quickly, describing this process of automaticity as "getting faster" and presented it as evidence of improvement. Increases in automatic responses were described with a sense of pride and appeared to provide a source of confidence for students.

> *You don't spend like a lot of time on it and stuff, you're able to like do it quickly and then have the answer.*
> *(Year 8 student, mid-intervention interview)*

> *Because you're improving your time, and then if like… you're doing a test, and there's lots of answers, you can just finish it. And you won't have to worry, because you'll know all the answers.*
> *(Year 4 student, mid-intervention interview)*

The pairing of students of similar ability levels in the regular QuickSmart sessions cultivated a sense of competition. Competition, particularly when tied to sound pedagogical practices, has been found to positively influence students' attitudes towards mathematics (Van Eck, 2006). Such a change in attitude brought about by competition was described by both students and QuickSmart instructors in the interviews.

> *Because you're doing it with a partner, you want to beat the partner*
> *(Year 8 student, mid-intervention interview)*

> *I think this gives the children who are not up there in the classroom that opportunity to be up there is this classroom in the sense that they can compete against themselves and improve their goals. But they have a bit of nice competition with their partner as well*
> *(QuickSmart Instructor, Secondary, post-intervention interview)*

Program feasibility

Classroom transition

According to the QuickSmart program design, students are required to attend three 30-minute QuickSmart lessons each week. Teachers in the Primary school context noted some challenges in transitioning between their classrooms and QuickSmart. A concern reported by students and teachers was the need for classroom teachers to support students to catch up on whatever they had missed during this time. One teacher spoke of some self-directed strategies she had tried to work on with her students, in order to facilitate their smoother return to the classroom from QuicksSmart:

> *We told them that you need to be a little bit more responsible and show some initiative, see if you can look around and see what's happening first*
> *(Year 4 teacher, post-intervention interview)*

In general, concerns such as these were counterbalanced by the more positive outcomes teachers associated with the program, as this school executive member comments:

> *Class teachers, I don't think there's any - I've never heard them say really bad things. They might have said maybe sometimes it's a pain if a child goes out of class but I think they're seeing the value of the program and they believe in it so much so that it doesn't bother them at all when students are coming and going.*
> *(QuickSmart Instructor, Primary, post-intervention interview)*

In general, Primary school students did not share any concerns about being withdrawn from their classroom to attend QuickSmart. As the year progressed, however, Primary students noted that they sometimes had to miss out on fun activities in the classroom and so sometimes they did not want to go to QuickSmart Sessions.

> *It's sometimes fun but mostly it's a bit annoying sometimes because when we're doing something fun, you have to stop and then you have to go to QuickSmart.*
> *(Year 4 student, mid-intervention interview)*

Some of the QuickSmart Instructors indicated that if students did not want to go to their QuickSmart session at that time, they would not put pressure on them to do so. Students' reticence to attend QuickSmart, however, can reduce the fidelity with which the program can be implemented in the school and the exposure for individual students.

Issues related to withdrawing students from classes were substantially more challenging in Secondary schools. A number of Secondary students reported that missing out on lessons that were seen as important was a source of frustration and anxiety. This anxiety increased towards the end of each term as class work focused more specifically on assessments. Anxiety with regards to assessment performance is a known factor influencing performance itself, particularly in mathematics (Meece, Wigfield & Eccles, 1990). One student, in particular, withdrew from the program as she found missing other classes to be too stressful.

> *So, I was getting taken out of class, as all the other girls did, but I was just finding it hard to - I was falling a bit behind in all my other classwork and all that. So I was just finding it really hard and I was stressing out and all that.*
> *(Year 8 student, post-intervention interview)*

> *The hardest thing is just coming out of class.*
> *(Year 8 student, post-intervention interview)*

QuickSmart Instructors also noted difficulties associated with ensuring that students attended all of their sessions. They reported that while the expectation was that Secondary students would attend scheduled QuickSmart sessions, they spent a substantial amount of time finding and reminding students, which limited the time that could be spent completing required activities.

> *When you've got to go and chase them, when it's only half an hour... you're given five minutes to get here. Then one of the partners are here, [but] they're not here. Then you've got to go out and get them… your time's gone. You want them to get back to class for that last half an hour.*
> *(QuickSmart Instructor, Secondary, post-intervention interview)*

I think QuickSmart is great in the Primary years. You've got to get them then. I think in high school it's just hard to get them onboard, because "what do my friends think?", "It's taking time out", "I've got to be in class to do this", or "my assessment, we're doing assessment work", or you know we've got - they want to be in class so they don't miss out. Because these kids are already sort of behind.
(QuickSmart Instructor, Secondary, post-intervention interview)

QuickSmart Instructors reported that they tried to minimise the impact of attending the program on students' learning in other curriculum areas. Instructors and school-level coordinators in Secondary schools, in particular, reported that they faced challenges in timetabling sessions to ensure that students were not missing parts of their mathematics class every time they attended the program, thereby sharing the load of attendance for students across Key Learning Areas:

They've done it on a rotating cycle, so they're not trying to miss too much of one subject but at the end of the day they're still getting withdrawn from the subject and then having to catch up on that subject and they're already girls who are already struggling.
(QuickSmart Instructor, Secondary, post-intervention interview)

Some QuickSmart Instructors in Secondary schools offered sessions before or after school hours to minimise the impact on student learning and anxiety. Instructors reported, however, that this strategy was not sustainable long-term as it placed additional pressure on students' families to ensure that they were at school for their sessions and it was a source of additional cost for the school to pay Instructors to deliver the program outside of regular hours.

Another strategy implemented in one of the Secondary schools to encourage student attendance and engagement, which was recommended in the Instructor training, was a merit system to reward those who were engaged in the program.

We learned this at the other workshop about like a merit system, and you know, [students] who are making an effort with attendance or make really big improvements, we've got an online like merit system, and just a teacher feedback - so I've been using a bit more of that, and the [students] seem to like that. Sometimes I'll say to them, "we don't have to play the set games, can you think of a different type?" We kind of meld the games together, just to kind of brighten it up a little bit.
(QuickSmart Instructor, Secondary, post-intervention interview)

Social consequences

An additional issue in the Secondary school context, described by teachers, was the social implications of withdrawing students from classrooms to attend QuickSmart as some students did not wish to attend and did not want to attract extra attention to themselves in front of their peers:

The other one has been negative. Just doesn't want to go and doesn't like being brought off class I guess as well which draws attention.
(Secondary Maths Teacher, post-intervention interview)

Students did not report any specific negative responses from their peers related to their attendance at the QuickSmart sessions. Secondary students, however, were far less likely than their Primary peers to openly discuss their involvement in the QuickSmart process.

4.2 Formative findings

What are the necessary conditions for success of the intervention?

The QuickSmart intervention did not demonstrate a significant overall effect on mathematics achievement. Of the two cohorts, the Primary intervention cohort received a greater proportion of sessions, and displayed a positive effect on student efficacy levels, and indicated a stronger effect within analysis adjusting for exposure. The difference in the conditions within each cohort, we believe, gives some insight into the conditions required to maximise the effects of the QuickSmart intervention. The differences in the conditions created within the Primary cohort, in comparison to the Secondary cohort, are outlined below. In comparison to the Secondary cohort, the Primary cohort displayed:

- Logistics management that enabled a greater proportion of students to receive more QuickSmart sessions across the intervention period;
- A far greater proportion of instructors who had undertaken the higher levels (Levels 2 & 3) of QuickSmart Instructor training (Primary = 62%, Secondary = 5.3%);
- A far more experienced cohort of instructors (>70% with more than three years of experience);
- Greater levels of mathematics teaching efficacy and outcome expectancy among instructors (likely as a consequence of training and experience); and
- Teachers who appeared to adapt to the challenge of transitioning between QuickSmart and ongoing classroom activities.

This evaluation indicated that Primary school structures were more flexible in supporting the needs of students and classroom teachers in order to minimise the potential disruption and support fidelity of implementation of the QuickSmart program. Generally, the QuickSmart Instructors in participating Primary schools had undertaken higher levels of training and had greater experience in delivering the program.

Due to the nature of Primary classrooms, teachers were more likely to be aware of when students were attending QuickSmart settings and had greater flexibility to ensure that they were able to complete any work that would contribute to assessment. All of these conditions reduced the potential for student anxiety and provided Primary students with access to a greater volume of QuickSmart sessions across the intervention period.

Timetabling, transitions from classroom, and program adherence appeared to be more complex in Secondary school settings.

Improved intervention delivery

The major barrier to the implementation of the QuickSmart program was the lack of consistent access to the intervention, which resulted in very few students in the evaluation achieving completion of the full 90-session, 30-week program. The intervention volume, or number of QuickSmart sessions provided, was substantially higher in Primary school settings and QuickSmart Instructors had received higher levels of training and had greater experience. As such, it is recommended that, in the initial phases of establishing QuickSmart at a school, schools seek the assistance of an experienced QuickSmart instructor to advise on the most effective methods for managing the logistics of the QuickSmart program. It is also recommended that schools utilise the available mechanisms for seeking assistance and advice from the SiMERR team at UNE regarding effective implementation.

The withdrawal from regular classes to attend the program placed additional responsibility on classroom teachers and students to ensure that QuickSmart students did not fall behind in other areas of their learning. One of the key issues faced by teachers is that those students who are involved in QuickSmart are often at risk of falling behind in other curriculum areas. Strategic timetabling of sessions is critical to the effective implementation of the program. In Secondary schools, this timetabling would require ongoing coordination with classroom teachers to ensure that students did not worry about missing necessary information and are supported in transitions between their classroom and QuickSmart sessions.

Higher levels of training and experience appear to be linked to higher quality delivery of the QuickSmart intervention. It is recommended that schools seek out experienced TAs or teachers trained as QuickSmart instructors to assist in the initial phases of QuickSmart delivery at a school whilst additional instructors are moving through the phases of training and gaining experience in the delivery of the intervention. Schools that were implementing the program for the first time as part of this evaluation regularly contacted the evaluation team with questions, reporting their need for greater access to support in implementing the program.

We also advise that schools who are buying the program make every effort to closely follow the process developed by the SiMERR team. Having an experienced staff member to oversee regular fidelity checking within a school, or having a central person oversee fidelity at multiple schools, would strengthen the delivery of the intervention. Whilst we acknowledge that this does impose an additional cost for the program, the value of standardising delivery and ensuring quality delivery and assistance to less experienced instructors is seen as a worthwhile investment, given the size of the investment already required by schools to undertake the program.

5 Conclusion

5.1 Interpretation

This trial was designed to test whether the gains in mathematics achievement attributed to students randomly allocated to the QuickSmart intervention were significantly greater than the gains made by students who were receiving regular classroom mathematics instruction. Both of these groups were expected to make gains in mathematics achievement, with the QuickSmart group expected to achieve greater gains. Designed as an effectiveness trial, this evaluation sought to replicate how QuickSmart is typically undertaken within a school setting.

When comparing the average exposure to the program for this trial against those reported from 2013 – 2016 by SiMERR (Pegg et al., 2013, 2014, 2015, 2016), trial average exposure was marginally greater among the Year 4 cohort (Trial = 73%; SiMERR = 66%), and marginally lower among the Year-8 cohort (Trial = 49%; SiMERR = 55%). Under these conditions, this trial produced no evidence that the QuickSmart program led to significant average gains in mathematics achievement beyond those produced by participation in regular classroom-based mathematics instruction.

When comparing the findings of this trial against the average reported program effects (Pegg et al., 2013, 2014, 2015, 2016) for QuickSmart participants only (independent of control group effects), the average effect on achievement in this trial was greater than reported among the Year-4 cohort (Trial d = 1.52; 95% CI: 1.08 – 1.91; SiMERR d = 0.94; 95% CI: 0.78 – 1.10), and mariginally lower that reported among the Year-8 cohort (Trial d = 0.57; 95% CI: 0.23 – 0.91; SiMERR d = 0.64; 95% CI: 0.53 – 0.75). It is suggested that the difference in results is likely a product of compliance, with the Year-4 group in this trial above SiMERR average reported compliance levels, and the Year-8 cohort slightly below reported levels. These results are comparable to those reported between 2013 and 2016 for QuickSmart participants, however under the conditions of an effectiveness trial involving a randomised control group, these achievement effects display no significant difference to those receiving regular classroom instruction.

Primary schools in our evaluation appeared better able to implement the QuickSmart intervention, with students able to participate in a greater volume of QuickSmart sessions on average than the Secondary cohort. QuickSmart Instructors in Primary schools were, on average more qualified and had more years of QuickSmart experience, perhaps leading to increased efficiency within Primary schools. Pragmatically, QuickSmart would appear to fit logistically within a Primary school setting, with a single classroom teacher able to oversee reminders for students, rather than having to facilitate logistics with a larger number of teachers in the Secondary setting. Our evaluation suggests that timetabling of QuickSmart sessions and classroom transitions were reported to be more difficult in Secondary school settings, and that highly trained QuickSmart Instructors could improve the level of implementation within schools (both Primary and Secondary) and support students to attend the desired number of QuickSmart sessions within the year.

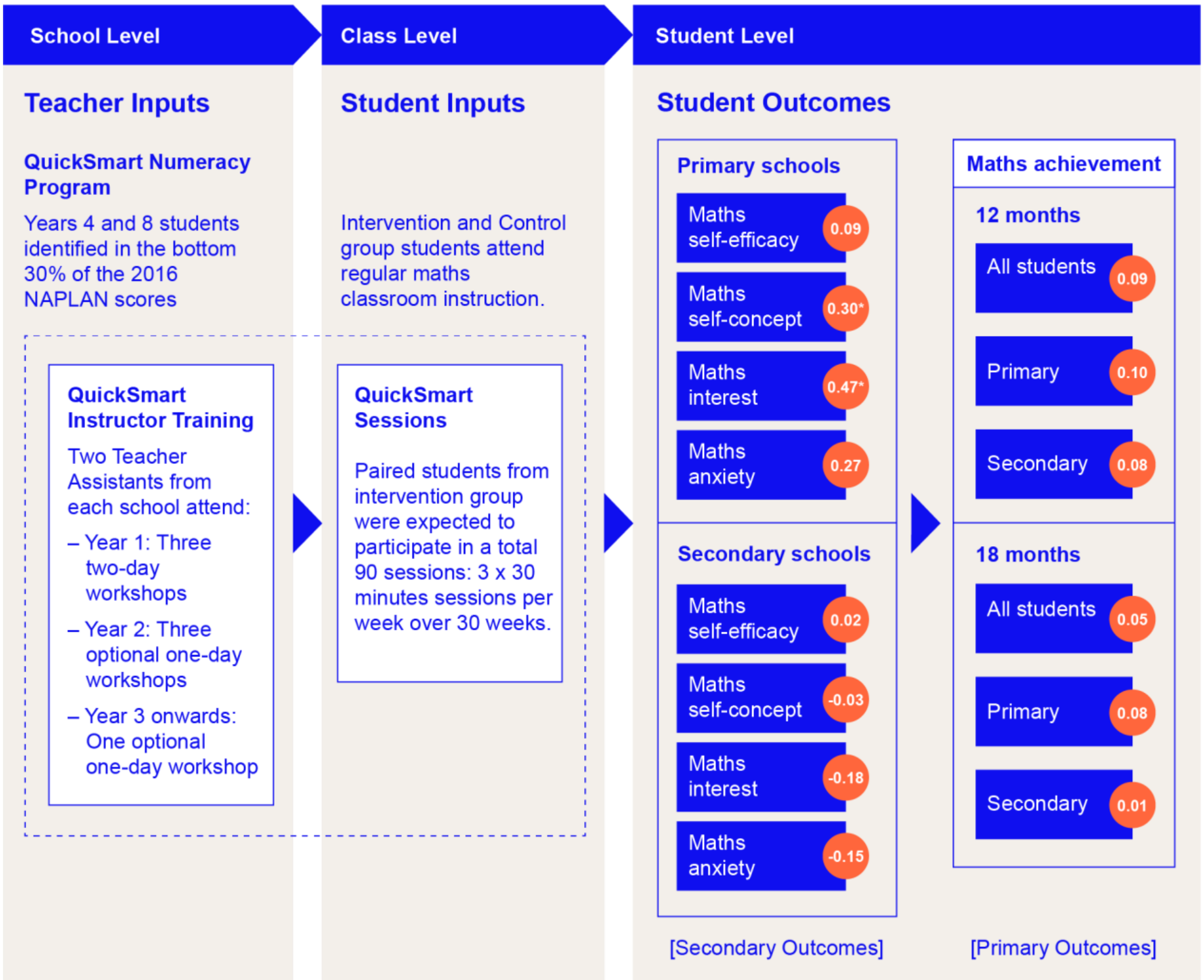
Primary school students involved in the QuickSmart program displayed significant gains for secondary outcomes, specifically cognitive and affective outcomes of mathematics self-concept and mathematics interest. Mathematics teachers within both cohorts reported that students involved in QuickSmart gained some confidence in responding to questions in their mathematics class, however any translation of this increased engagement to positive affective outcomes was limited to the Primary cohort. It is evident from the process evaluation that Instructors, teachers and students valued the intervention and believe that it had a positive effect on students' confidence and engagement in mathematics.

The major issues experienced in QuickSmart schools highlighted transitioning between the classroom and QuickSmart instruction without major disruption, and the reduction of volume for learning in other subject areas replaced by QuickSmart instruction. Schools indicated a desire for greater support to implement the program. We also recommend increased coordination with Secondary mathematics and Primary classroom teachers to effectively implement the QuickSmart program and support the achievement of the greatest possible return on investment.

Key Conclusions
3. In this trial, QuickSmart did not have an additional impact on maths achievement compared to regular classroom instruction and support. There was a small positive gain, equivalent to one month's additional learning, however the trial was not commissioned to detect this level of difference[4] meaning the difference was not statistically significant.
4. When models were adjusted for intervention exposure, there was a small increase in the effect on student achievement (indicating that exposure levels have some effect on outcomes), however this effect was not statistically significant.
5. Sub-group analysis displayed a small but not statistically significant positive effect for Primary students. The gain was equivalent to one month's additional learning. There was no additional effect for Secondary students.
6. Schools faced challenges achieving the prescribed program exposure of 90 sessions within 30 weeks. Primary students, on average, received 73% (or 66 sessions) of QuickSmart's prescribed 90 sessions over 30 weeks, while Secondary students received 49% (or 44 sessions). Only 35% of Primary students and 4% of Secondary students received more than 75% (or 67 sessions) of the prescribed QuickSmart sessions.
7. Sound implementation of QuickSmart appeared more feasible within Primary schools than Secondary schools. Both settings struggled with transitions into and out of the classroom, and concern about the subject matter students were missing out on as a result of QuickSmart was expressed across Primary and Secondary settings.
8. Primary teachers were positive about QuickSmart and reported that it appeared to help students gain more confidence participating in their maths classrooms. QuickSmart had a statistically significant positive impact on Primary students' maths self-concept (effect size g = 0.30) and interest in maths (effect size g = 0.47), however there was no evidence of impact on self-efficacy (effect size g = 0.09). There were no statistically significant intervention effects on Secondary students' cognitive and affective outcomes.

[4] This trial was powered to achieve a Minimum Detectable Effect Size (MDES) of 0.24 at randomisation, which meets the high padlock rating criteria for MDES of <0.3.

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



*indicates statistically significant effect (p<0.05)

5.2 Limitations

The main limitation of this study is that teachers, particularly Primary school teachers, knew which students were allocated to each of the study conditions. The transparency of those involved in the intervention and control conditions produced potential motivation bias as teachers may exert more effort in teaching the students in the control condition to ensure they don't fall behind without the additional QuickSmart intervention. Classroom observations were not undertaken to assess if teachers gave either condition more attention. However, given the volume of instruction undertaken by classroom teachers on a daily basis, and the clustered design of the trial, the potential for this to affect the results is considered minor.

Additionally, we were unable to control the classes that students were allocated to between post-intervention and 6-month follow-up assessment. As students only received 8-weeks of instruction within a new class during Term 1 of the 6-month post-intervention period, and some students may have remained in the same class grouping in the new year, the effect of the bias introduced through different instructional quality is not seen as detrimental to the effects of the intervention.

Among the instructors in this trial, there was a broad range of QuickSmart training and experience. There is no information recorded or reported by SiMERR regarding the training and experience of QuickSmart instructors undertaking programs in schools, and the balance of training and experience in this trial was common to the broader scale is unknown. Whilst it appears that the level of experience did not have an effect on the level of exposure among the schools involved, this must still be accepted as a limitation in this trial.

5.3 Future research and publications

There is not enough evidence to claim with certainty that the QuickSmart program, as is currently implemented in schools, improves mathematics achievement over and above students receiving regular classroom tuition. This finding suggests the need for further trials to ensure better control of intervention delivery and test different modes of delivery for the QuickSmart intervention (e.g., highly trained instructors visiting a group of schools to standardise delivery quality).

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2011). *Guide to understanding ICSEA*. Sydney, Australia: Author. Retrieved from: http://docs.acara.edu.au/resources/Guide_to_understanding_ICSEA.pdf

Abel, M., Hannon, J., Mullineaux, D., & Beighle, A. (2011). Determination of step rate thresholds corresponding to physical activity intensity classifications in adults. Journal of Physical Activity & Health, 8(1), 45-51.

Baker, S., Gersten, R., & Lee, D.S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. Elementary School Journal, 10, 51-73.

Bandura, A. (1997), Self-Efficacy: the Exercise of Control, Freeman, New York.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions. Educational Evaluation and Policy Analysis, 29(1), 30-59. doi:10.3102/0162373707299550

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. Quantitative methods in psychology, 112(1), 155-159.

Deshler, D., Mellard, D. F., Tollefson, J. M., & Byrd, S. E. (2005). Research topics in responsiveness to intervention. Journal of Learning Disabilities, 38(6), 483-484.

Donner, A., & Klar, N. (2000). Design and Analysis of Cluster Randomization Trials in Health Research: Arnold.

Dreyhaupt, J., Mayer, B., Keis, O., Öchsner, W., & Muche, R. (2017). Cluster-randomized Studies in Educational Research: Principles and Methodological Aspects. GMS journal for medical education, 34(2), Doc26-Doc26. doi:10.3205/zma001103

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics, 100*(4), 194–202.

Ferla, J., Valke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences, 19*, 499–505.

Fuchs, D. & Fuchs, L. S. (2001). Principles for the prevention and intervention of mathematical difficulties. Learning Disabilities Research and Practice, 16, 85-95.

Lamb, S., & Fullarton, S. (2001) Classroom And School Factors Affecting Mathematics Achievement: a Comparative Study of the US and Australia Using TIMSS. In. http://research.acer.edu.au/timss_monographs/10.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 520-540.

Marston, D. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities patterns. Journal of Learning Disabilities, 38(6), 539-544.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445-463.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of educational psychology*, *82*(1), 60.

National Assessment Program – Literacy and Numeracy (NAPLAN). (2018). *National protocols for test administration 2018*. Retrieved from: https://www.nap.edu.au/about

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

Pegg, J., & Graham, L. (2013). A three-level intervention pedagogy to enhance the academic achievement of Indigenous students: Evidence from QuickSmart Mathematics research relevant to Indigenous populations: Evidence-based practice. In R. Jorgenson, P. Sullivan & P. Grootenboer (Eds.), *Pedagogies to enhance learning for Indigenous students* (pp. 123-138). Singapore: Springer.

Pegg, J., Horarik, S., McDermott, A., & Billings, J. (2013). Annual Numeracy Program Report 2013. Retrieved from Armidale NSW: https://simerr.une.edu.au/quicksmart/pages/qsresearchevidence.php

Pegg, J., Horarik, S., McDermott, A., & Billings, J. (2014). Annual Numeracy Program Report 2014. Retrieved from Armidale NSW: https://simerr.une.edu.au/quicksmart/pages/qsresearchevidence.php

Pegg, J., Horarik, S., McDermott, A., & Billings, J. (2015). Annual Numeracy Program Report 2015. Retrieved from Armidale NSW: https://simerr.une.edu.au/quicksmart/pages/qsresearchevidence.php

Pegg, J., Horarik, S., McDermott, A., & Billings, J. (2016). Annual Numeracy Program Report 2016. Retrieved from Armidale NSW: https://simerr.une.edu.au/quicksmart/pages/qsresearchevidence.php

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, *93*(2), 346.

Rowe, K., Stephanou, A., & Urbach, D. (2006). Effective teaching and learning practices initiative for students with learning difficulties. Report to the Australian Government Department of Education, Science and Training (DEST). Canberra: Australian Government Printery.

SiMERR National Research Centre (2017). QuickSmart Annual Numeracy Program Report 2016. UNE, Armidale.

Westwood, P. (2007). What teachers need to know about numeracy. Melbourne: ACER.

Wigfield, A. and J.S. Eccles (2000), “Expectancy - value theory of motivation”, Contemporary Educational Psychology, 25, pp. 68-81.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Changes in children’s competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology, 89*, 451–469.

Van Eck, R. (2006). The effect of contextual pedagogical advisement and competition on middle-school students' attitude toward mathematics using a computer-based simulation game. *Journal of Computers in Mathematics and Science Teaching*, 25(2), 165-195.

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 for Evidence for Learning's trials is based on the translation of effect size to months' progress established by the Education Endowment Fund (EEF) (Higgins et al., 2013). Source: educationendowmentfoundation.org.uk/help/projects/the-eefs-months-progress-measure

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 of trial findings

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

Appendix C: QuickSmart Fidelity Checklist

The 13-item checklist used by the evaluators to check fidelity of the QuickSmart intervention is provided below.

QuickSmart research project – session checklist

Please place a tick in each row for any items that WERE completed:

QuickSmart Instructor's Name: ____________ Student's Name: ____________

School: ____________

	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:	Date:
Attendance (please note if student is absent on usual QS days)																		
QuickSmart week																		
QS instructor initials																		
Focus Facts																		
Flash Cards																		
Completion																		
Correction																		
Student's graph																		
Feedback																		
Speed Sheet																		
Completion																		
Correction																		
Student's Summary																		
Independent Worksheet																		
Completion																		
Feedback																		
OZCAAS																		
Completion																		
View graph of results																		
Discussion																		
Game																		

Appendix D: Student questionnaire – Year 4 (baseline for example)

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

QuickSmart Numeracy Year 4 Student Questionaire

* 1. What is your full name? (First name and last name)

* 2. Which school do you attend?

3. How confident do you feel about having to do the following Mathematics tasks?

	Very confident	Confident	Not very confident	Not at all confident
Adding two numbers in the hundreds.	◯	◯	◯	◯
Subtracting two numbers in the hundreds.	◯	◯	◯	◯
Multiplying any number by 2.	◯	◯	◯	◯
Multiplying any number by 7.	◯	◯	◯	◯
Understanding graphs presented in newspapers.	◯	◯	◯	◯
Changing measuring units from centimetres to metres	◯	◯	◯	◯
Identifying shapes by the number of sides they have (for example a triangle or a hexagon)	◯	◯	◯	◯
Calculating the decimal value of a simple fraction like 3/4	◯	◯	◯	◯

1

4. Thinking about studying Mathematics: to what extent do you agree with the following statements?

	Strongly agree	Agree	Disagree	Strongly disagree
I am just not good at Mathematics.	○	○	○	○
I get good marks in Mathematics.	○	○	○	○
I learn Mathematics quickly.	○	○	○	○
I have always believed that Mathematics is one of my best subjects.	○	○	○	○
In my Mathematics class, I understand even the most difficult work.	○	○	○	○

5. Thinking about your views on Mathematics: to what extent do you agree with the following statements?

	Strongly agree	Agree	Disagree	Strongly disagree
I enjoy reading about Mathematics.	○	○	○	○
I look forward to my Mathematics lessons.	○	○	○	○
I do Mathematics because I enjoy it.	○	○	○	○
I am interested in the things I learn in Mathematics.	○	○	○	○

6. Thinking about studying Mathematics: to what extent do you agree with the following statements?

	Strongly agree	Agree	Disagree	Strongly disagree
I often worry that it will be difficult for me in Mathematics classes.	○	○	○	○
I get very stressed when I have to do Mathematics homework.	○	○	○	○
I get very nervous doing Mathematics problems.	○	○	○	○
I feel helpless when doing a Mathematics problem.	○	○	○	○
I worry that I will get poor marks in Mathematics.	○	○	○	○

2

Appendix E: Student questionnaire – Year 8 (baseline for example)

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

QuickSmart Numeracy Year 8 Student Questionaire

* 1. What is your full name? (First name and last name)

* 2. Which school do you attend?

3. How confident do you feel about having to do the following Mathematics tasks?

	Very confident			Not at all confident
Using a train timetable to work out how long it would take to get from one place to another.				
Calculating how much cheaper a TV would be after a 30% discount.				
Calculating how many square metres of tiles you need to cover a floor.				
Understanding graphs presented in newspapers.				
Solving an equation like 3x+ 5= 17				
Finding the actual distance between two places on a map with a 1:10,000 scale.				
Solving an equation like 2(x+ 3)= (x+ 3)(x−3)				
Calculating the petrol consumption rate of a car.				

1

4. Thinking about studying Mathematics: to what extent do you agree with the following statements?

	Strongly agree			Strongly disagree
I am just not good at Mathematics.	○	○	○	○
I get good marks in Mathematics.	○	○	○	○
I learn Mathematics quickly.	○	○	○	○
I have always believed that Mathematics is one of my best subjects.	○	○	○	○
In my Mathematics class, I understand even the most difficult work.	○	○	○	○

5. Thinking about your views on Mathematics: to what extent do you agree with the following statements?

	Strongly agree			Strongly disagree
I enjoy reading about Mathematics.	○	○	○	○
I look forward to my Mathematics lessons.	○	○	○	○
I do Mathematics because I enjoy it.	○	○	○	○
I am interested in the things I learn in Mathematics.	○	○	○	○

6. Thinking about studying Mathematics: to what extent do you agree with the following statements?

	Strongly agree			Strongly disagree
I often worry that it will be difficult for me in Mathematics classes.	○	○	○	○
I get very tense when I have to do Mathematics homework.	○	○	○	○
I get very nervous doing Mathematics problems.	○	○	○	○
I feel helpless when doing a Mathematics problem.	○	○	○	○
I worry that I will get poor marks in Mathematics.	○	○	○	○

2

Appendix F: Instructor questionnaire

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

QuickSmart Numeracy Teacher Questionnaire

* 1. What is your full name? (first and last name)

* 2. Where do you teach?

3. Please indicate the degree to which you agree or disagree with each statement below.

	Strongly agree	Agree	Uncertain	Disagree	Strongly Disagree
When a student does better than usual in mathematics, it is often because the teacher exerted a little extra effort.	○	○	○	○	○
I will continually find better ways to teach mathematics.	○	○	○	○	○
Even if I try very hard, I do not teach mathematics as well as I do most subjects.	○	○	○	○	○
When the mathematics grades of students improve, it is often due to their teacher having found a more effective teaching approach.	○	○	○	○	○
I know the steps necessary to teach mathematics concepts effectively.	○	○	○	○	○
I am not very effective in monitoring mathematics activities.	○	○	○	○	○

	Strongly agree	Agree	Uncertain	Disagree	Strongly Disagree
If students are underachieving in mathematics, it is most likely due to ineffective mathematics teaching.	◯	◯	◯	◯	◯
I generally teach mathematics ineffectively.	◯	◯	◯	◯	◯
The inadequacy of a student's mathematics background can be overcome by good teaching.	◯	◯	◯	◯	◯
The low mathematics achievement of some students cannot generally be blamed on their teachers.	◯	◯	◯	◯	◯
When a low-achieving child progresses in mathematics, it is usually due to extra attention given by the teacher.	◯	◯	◯	◯	◯
I understand mathematics concepts well enough to be effective in teaching mathematics.	◯	◯	◯	◯	◯
Increased effort in mathematics teaching produces little change in some students' mathematics achievement.	◯	◯	◯	◯	◯
The teacher is generally responsible for the achievement of students in mathematics.	◯	◯	◯	◯	◯
Students' achievement in mathematics is directly related to their teacher's effectiveness in mathematics teaching.	◯	◯	◯	◯	◯

	Strongly agree	Agree	Uncertain	Disagree	Strongly Disagree
If parents comment that their child is showing more interest in mathematics at school, it is probably due to the performance of the child's teacher.	◯	◯	◯	◯	◯
I find it difficult to use manipulatives to explain to students why mathematics works.	◯	◯	◯	◯	◯
I am typically able to answer students' mathematics questions.	◯	◯	◯	◯	◯
I wonder if I have the necessary skills to teach mathematics.	◯	◯	◯	◯	◯
Given a choice, I would not invite the principal to evaluate my mathematics teaching.	◯	◯	◯	◯	◯
When a student has difficulty understanding a mathematics concept, I am usually at a loss as to how to help the student understand it better.	◯	◯	◯	◯	◯
When teaching mathematics, I usually welcome student questions.	◯	◯	◯	◯	◯
I do not know what to do to turn students on to mathematics.	◯	◯	◯	◯	◯