2021

Working with practitioners and learning trajectories: Sharpening the focus on mathematical reasoning in grades 5-9

Dianne Siemon

Lorraine Day

Rosemary Callingham

Marj Horne

This conference paper was originally published as:
Siemon, D., Day, L., Callingham, R., & Horne, M. (2021). Working with practitioners and learning trajectories: Sharpening the focus on mathematical reasoning in grades 5-9. *AERA 2021 symposium on Partnering with Practicing Teachers to Strengthen Inclusive Instruction Using Learning Trajectories*.

Working with practitioners and learning trajectories: Sharpening the focus on mathematical reasoning in grades 5-9[1]

Dianne Siemon, RMIT University

Lorraine Day, University of Notre Dame Australia

Rosemary Callingham, University of Tasmania

Marj Horne, RMIT University

This paper will report on the role of practitioners in a recent Australian study that developed empirically based learning and assessment frameworks (i.e. learning trajectories) for algebraic, geometrical, and statistical reasoning in the middle years of schooling. To understand the nature of the teachers' role, the paper begins with a description of what is meant by 'curriculum' in Australia and the implications of this for teacher decision making and planning. We then provide a rationale for the study and a brief description of the methodology before illustrating how teachers were involved in the iterative research design through task development and the trial and refinement of partial credit scoring rubrics. The paper concludes by describing the development of targeted teaching advice and considering some of the challenges involved in dissemination.

Understanding the context

The term 'curriculum' has different meanings in different countries and contexts. In the United States for instance, the term generally refers to an ordered sequence of instructional materials that specify what students are expected to learn on a daily or weekly basis and provide detailed lesson plans, learning activities, teaching resources, and the assessments needed to teach and evaluate learning (e.g. Clements, 2007). Informed by State-based year level standards such as the *Common Core State Standards for Mathematics* [CCSSM], the development of these materials often involves a partnership between commercial publishers and researchers in mathematics education. When a mathematics curriculum is adopted by the relevant authorities, there may be a period of induction and professional development but thereafter teachers are expected to implement the curriculum as intended.

The Common Core State Standards Mathematics are not a curriculum in this sense. They do not 'tell teachers what to teach'. The Standards aim to provide "clarity and specificity rather than broad general statements" and they value conceptual understanding and procedural fluency alongside problem solving, reasoning, and mathematical practices. While they recognise societal expectations, the Standards are based on "research-based learning progressions detailing what is known today about how students' mathematical knowledge, skill, and understanding develop over time" (http://www.corestandards.org/Math/).

In Australia, 'curriculum' refers to a nationally agreed document that outlines what is to be taught and the quality of learning expected of young people as they progress through school (Australian Curriculum Assessment & Reporting Authority [ACARA], 2016). As such, it inevitably represents a compromise between the State and Territory authorities who retain responsibility for the provision of education and what research suggests might be optimal learning pathways.

[1] Paper presented at the AERA 2021 symposium on *Partnering with Practicing Teachers to Strengthen Inclusive Instruction Using Learning Trajectories*

The *Australian F-10 Curriculum: Mathematics* [AC:M] (ACARA, 2016) is framed in terms of three content strands and four proficiency strands. The content strands are number and algebra, measurement and geometry and statistics and probability. They describe in broad terms the behaviours expected of students at each year level of primary and secondary school from Foundation (first year of primary school) to Year 10. To demonstrate the nature of these, a sample of the Year 4 Number and Algebra content descriptors is given below.

> Recognise, represent and order numbers to at least tens of thousands (ACMNA072)
>
> Recall multiplication facts up to 10 × 10 and related division facts (ACMNA075)
>
> Count by quarters halves and thirds, including with mixed numerals. Locate and represent these fractions on a number line (ACMNA078)
>
> Find unknown quantities in number sentences involving addition and subtraction and identify equivalent number sentences involving addition and subtraction (ACMNA083) (ACARA, 2016)

The proficiency strands are understanding, fluency, problem solving, and mathematical reasoning. While these are intended to "reinforce the significance of working mathematically within the content and describe how the content is explored or developed" (ACARA, 2016), the year level descriptions of these tend to echo curriculum expectations rather than higher order thinking, complex problem solving, or reasoning as shown by the Year 7 level description given below.

> *understanding* includes describing patterns in uses of indices with whole numbers, recognising equivalences between fractions, decimals, percentages and ratios, plotting points on the Cartesian plane, identifying angles formed by a transversal crossing a pair of lines, and connecting the laws and properties of numbers to algebraic terms and expressions
>
> *fluency* includes calculating accurately with integers, representing fractions and decimals in various ways, investigating best buys, finding measures of central tendency and calculating areas of shapes and volumes of prisms
>
> *problem-solving* includes formulating and solving authentic problems using numbers and measurements, working with transformations and identifying symmetry, calculating angles and interpreting sets of data collected through chance experiments
>
> *mathematical reasoning* includes applying the number laws to calculations, applying known geometric facts to draw conclusions about shapes, applying an understanding of ratio and interpreting data displays. (ACARA, 2016)

These examples of the content descriptors and year level proficiency descriptions from the AC:M stand in marked contrast to the CCSSM, which, in addition to being based on evidenced-based learning progressions and providing "clarity and specificity" at each Year level, also include mathematical practices such as "reason abstractly and quantitively" and "construct viable arguments and critique the reasoning of others".

As indicated above, while the Standards speak to teachers they are also used to inform the development of detailed day-to-day instructional materials by large teams of researchers and practitioners for adoption by education authorities at a State or District level. By contrast, while there is a range of commercially produced resources to support Australian teachers enact the AC:M on a day-to-day basis, these tend to be developed with little or no input from researchers or significant investment in time and personnel. One of the possible reasons for this is that a fundamental principle on which the AC:M was developed and agreed to by the States and Territories was the primacy of teacher decision making (Sullivan, 2012). In Australia, there is no history of or expectation that commercially produced resources will be endorsed, adopted or mandated by education authorities. As a result, Australian teachers have considerable discretion in deciding exactly what mathematics is taught, when, and how it is taught and assessed. Even where schools adopt a particular resource, teachers decide how much time to spend on a particular aspect of mathematics, where in the year it might be considered if at all, and what other resources might be useful for teaching and/or assessment purposes. A consequence of this and the way in which the AC:M was constructed is that there is considerable variation in the quality of teacher decision making and thereby considerable variation in the opportunities students have to learn important mathematics. While there are a number of other factors involved such as the significant number of out-of-field teachers teaching mathematics in Years 7 to 9 (Weldon, 2016), this suggests there is a need for an evidenced-based resource that would build a bridge between the very general content descriptors and proficiency descriptions of the AC:M and what was known to be important in deepening and progressing students' mathematics learning.

Learning progressions/trajectories

Identifying and building on what students know is widely regarded as essential to success in school mathematics (Masters, 2013; Wiliam, 2013). However, determining what is important, when, and identifying what students understand in relation to what is deemed to be important are by no means uncontested or straightforward endeavours (Siemon, 2019). In recent years attention has turned to the development of evidenced-based learning trajectories (or progressions) as a means of identifying what mathematics is important and how it is understood over time (Clements & Samara, 2004; Daro, Mosher, & Corcoran, 2011). But for this information to be useful to practitioners, it needs to be accompanied by accurate forms of assessment that locate where learners are in their learning journey and evidenced-based advice about where to go to next.

From our perspective a learning trajectory is an integrated, empirically based learning and assessment framework that is focused on the development of one or more big ideas in mathematics and the links between them (Siemon, 2019). Consistent with more familiar definitions of learning trajectories (e.g. Clements & Samara, 2004; Confrey, & Maloney, 2010), the framework incorporates an evidence-based learning progression that serves as a model of how students' thinking evolves over time, validated diagnostic assessment tools that identify where students are in relation to the learning progression, and targeted teaching advice to help teachers progress students' learning from one level/zone of the progression to the next.

One of the reasons for the focus on big ideas was to make learning trajectories "more comprehensible and useful to practitioners" (Baroody et al., 2004; p, 253) thereby increasing the likelihood that teachers would use them to make more informed decisions about the next steps in instruction than they might otherwise have made in the absence of that evidence (Wiliam, 2011).

Why focus on mathematical reasoning

Numerous industry and government sponsored reports pointed to the threat to Australia's economy posed by lack of access to a suitably qualified workforce with knowledge, skills and experience in science, technology, engineering and mathematics (i.e. STEM) (Australian Industry Group, 2015; Business Council of Australia, 2015; PricewaterhouseCooper, 2015; Office of the Chief Scientist, 2016). This is reflected in the sharp decline in the proportion of students undertaking higher level mathematics and science subjects in the final years of schooling (Australian Industry Group, 2015; Australian Mathematical Sciences Institute, 2020) and the significant decline in Australian students' performance on international assessments of mathematical literacy relative to other countries (Thomson, De Bortoli & Underwood, 2016; Thomson, Wernet, O'Grady & Rodriguez, 2016).

Earlier Australian research that found a seven-year range in mathematics achievement in every year level from Year 5 to Year 9, which was due almost entirely to the extent to which students had access to multiplicative thinking (Siemon & Virgona, 2001; Siemon et al., 2006), helps explain this situation, but it is not the only reason. Despite the endorsement of important mathematical practices such as problem-solving and mathematical reasoning in the AC:M, the nature of the content descriptors and the brevity of the proficiency descriptions present mathematics as a set of disconnected topics and skills to be demonstrated and practised rather than explored, discussed and connected (Shield & Dole, 2013; Siemon, Bleckly & Neal, 2012; Sullivan, 2011). This in turn, prompts relatively narrow forms of assessment that value performance over mastery (Dweck & Leggett, 1998; Sullivan, 2011) and devalue problem solving and mathematical reasoning with the result that, along with these important mathematical practices, 'big ideas' such as multiplicative thinking and proportional reasoning are rarely given the attention they deserve (Siemon, 2017).

A focus on mathematical reasoning was needed to provide the sort of evidence and resources needed to support a significant and sustained change in practice away from low-complexity, procedural exercises (e.g. Vincent & Stacey, 2008) to teaching based on a deeper understanding of the big idea and the connections between them (Siemon et al., 2012; Sullivan, 2011). Defined broadly in the ACM as a "capacity for logical thought and actions", mathematical reasoning has a lot in common with mathematical problem solving, but it also relates to students' capacity to see beyond the particular to generalize and represent structural relationships, which are key aspects of further study in science technology, engineering and mathematics (i.e., STEM) related fields (Wai, Lubinski & Benbow, 2009).

Another reason for this focus was that despite the demonstrated efficacy of working with an evidenced-based learning and assessment framework for multiplicative thinking (e.g. Siemon et al., 2006a; Siemon, Banks, & Prasad, 2018), many teachers found it difficult to see the connection between multiplicative thinking and the mathematics content descriptors at their Year level. This is despite the fact that an analysis of the AC:M found that approximately 75% of the Year 8 curriculum required or assumed student access to multiplicative thinking (Siemon, 2013). This situation points to the critical importance of establishing an evidenced-based relationship between curriculum, instruction (i.e., teaching/pedagogy) and assessment (Black, Wilson, & Yao, 2011). However, it also points to the need to work closely with teachers as they are responsible for the day-to-day decisions about what mathematics is taught, when, and how it is taught and assessed.

This situation is not unique to Australia. Compared to the curricula of countries that do well on international assessments of mathematics achievement, the elementary and middle school mathematics curriculum in the United States has been "characterised as shallow, undemanding, and diffuse in content coverage" (Kilpatrick, Swafford & Findell, 2001, p. 4) leading to calls by these authors and others (e.g., Daro, Mosher, & Corcoran, 2011; Stacey, 2010; Sullivan, 2011) for a much greater focus on mathematical problem solving and reasoning.

This focus was both pragmatic - it had the potential to address the concerns of teachers who felt constrained from adopting a targeted teaching approach to multiplicative thinking by the perceived demands of the curriculum; and theoretical - it offered an opportunity to build probabilistic models of student learning in relation to mathematical reasoning that could be used to inform an evidenced-based approach to the teaching of algebra, geometry, and statistics in the middle years.

The *Reframing Mathematical Futures II* (RMFII) Project

The RMFII was a four-year research project funded by the Australian Government Department of Education and Training under the auspices of the Australian Mathematics and Science Partnership Programme. This built on two earlier projects, the *Scaffolding Numeracy in the Middle Years* (SNMY) project (Siemon et al., 2006) that established the evidenced-based learning and assessment framework for multiplicative thinking referred to above and the *Reframing Mathematical Futures-Priority* (RMF-P) (Siemon, 2016; Siemon, Banks, & Prassad, 2018) that explored the efficacy of using the SNMY formative assessment resources in secondary schools.

The aim of the RMFII project was to work with practitioners to build a sustainable, evidence-based, integrated learning and teaching resource to support the development of mathematical reasoning in Years 7 to 10 comprised of:

- evidence-based learning progressions in algebraic, statistical and spatial reasoning that can be used to inform teaching decisions and the choice of mathematics learning activities and resources by teachers and students.
- a range of validated, rich assessment tasks and scoring rubrics that can be used to identify what students know and understand in terms of the learning progressions, inform starting points for teaching and show learning over time (i.e. as pre and post tests);
- detailed teaching advice linked to the learning progressions that establish and consolidate learning at the level identified and introduce and develop the ideas and strategies needed to progress learning to the next level of the framework; and
- indicative resources to support the implementation of a targeted teaching approach in mixed ability classrooms.

For the purposes of the RMFII project, mathematical reasoning was defined in terms of three core elements:

- core knowledge needed to recognise, interpret, represent and analyse algebraic, spatial, statistical and probabilistic situations, and the relationships/connections between them;
- an ability to apply that knowledge in unfamiliar situations to solve problems, generate and test conjectures, make and defend generalisations; and
- a capacity to communicate reasoning and solution strategies in multiple ways (i.e., through diagrams, symbols, orally and in writing). (Siemon, 2016, p. 76)

A research team with expertise in one or more of the focus areas of mathematical reasoning and educational measurement was established to address the following research questions.

(i) To what extent can we develop rich tasks to accurately identify key points in the development of mathematical reasoning in the junior secondary years?

(ii) To what extent can we gather evidence about each student's achievements with respect to these key points to inform the development of a coherent learning and assessment framework for mathematical reasoning?

(iii) To what extent does working with the tasks and the knowledge they provide about student understanding assist teachers to improve student's mathematical performance at this level?

(iv) What strategies and/or teaching approaches are effective in scaffolding mathematical reasoning in the middle years?

(v) What are the key features of classroom organisation, culture and discourse needed to support/scaffold students' mathematical reasoning at this level?

With a view to disseminating and scaling up project outcomes to a wider professional audience (Cobb & Jackson, 2011), representatives from the Australian Association of Mathematics Teachers (AAMT) and State and Territory Departments of Education were invited to partner in the project. The Chief Executive Officer and the National Projects Manager from AAMT participated directly as members of the research team. State and Territory partners received all project information and were invited to participate in and contribute to the associated residential and online professional development sessions. They were also responsible for the identification of potential project schools.

Method

Given the ambitious nature of the project and the commitment to work with teachers, design-based research methods were seen to be most appropriate (e.g., Barab & Squire, 2004; Design-Based Research Collective, 2003). An important goal of design-based research is to "directly impact practice while advancing theory that will be of use to others" (Barab & Squire, p. 8). Design studies are typically interventionist and conducted in naturalistic settings to better understand the “messiness of real-world practice” (p. 3). They generally involve a multi-disciplinary team working with practitioners in successive iterations of “design, enactment, analysis, and redesign” (Design-Based Research Collective, 2003, p. 5) to develop “theories about both the process of learning and the means … to support that learning” (Cobb, Confrey, di Sessa, Lehrer, & Schauble, 2003, p. 10).

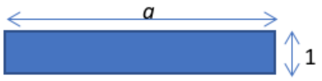
Thirty-two secondary schools from six of the seven Australian States or Territories participated in the RMFII project. A condition of funding was that the schools were located in lower socio-economic areas. A range of schools were invited by project partners to participate in the project. Where this was agreed and ethics approved, project funding was provided to support one teacher from each school (the Specialist) work with at least two other teachers in their school. From late 2014 to early 2018, approximately 80 teachers, and 3500 students in Years 7 to 10 were involved in the project. Residential professional learning opportunities were provided on an annual basis for the Specialists and at least one other teacher from each school where possible. Regular online professional learning sessions were provided for all involved or interested teachers. Project schools were visited at least twice a year by a member of the research team. An additional 1500 or so Year 5 to 10 students from a range of other schools (sourced through AAMT) participated in the trialing of the assessment tasks where insufficient data from project schools was available at the lower and upper ends of the hypothetical learning progressions and to generate a larger pool of responses from a broader range of schools.

Phase 1

This phase of the RMFII project involved an extensive literature review to identify the ‘big ideas’ in algebraic, geometrical, and statistical reasoning and inform the development of hypothetical learning progressions in each area. Research team members with specific expertise in one or more of the three areas were charged with identifying assessment tasks and partial credit rubrics that, where possible, could assess reasoning vertically (i.e., at different levels of complexity within the same hypothetical learning progression) and horizontally (i.e., at similar levels of difficulty across different hypothetical learning progressions). An example of one such task is the *Algebra Tiles Task*, the final version of which is shown in Figure 1 below. This task linked aspects of the hypothethetical learning progressions for geometrical reasoning (complex perimeter) and algebraic reasoning (use of symbolic text). In general, tasks were developed around a meaningful context (e.g. packaging a gift to be sent overseas, finding the best route in an emergency, or making sense of statistical claims made in relation to a school-wide survey), but in other circumstances tasks were set in decontextualized settings (e.g. explaining why a given relationship is true or false).

ATILP1

This tile has length of *a* units and width of 1 unit.



Without measuring, what is the perimeter of the tile?

ATILP2

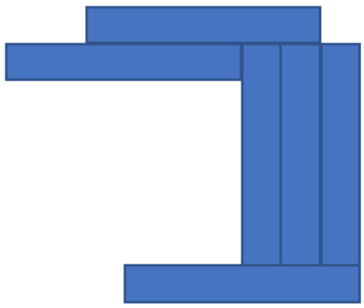
The shape below is made of the same tiles. Without measuring, what is the perimeter of the shape?

ATILP3

The shape on the right is made of the same tiles.

What is the perimeter of this shape?

Explain your reasoning.



ATILP1

SCORE	DESCRIPTION
0	No response or irrelevant response
1	Calculation based on numbers (e.g., assumes a = 5 or 6 cm) that shows an understanding of perimeter (e.g., 5 + 1 + 5 + 1 = 11 or 12 + 2 = 14)
2	Correct response but not in its simplest form (e.g., a + a + 1 + 1)
3	Correct, simplified response (2a + 2 or 2(a + 1))

ATILP2

SCORE	DESCRIPTION
0	No response or irrelevant response
1	Incorrect, but attempt made to solve for the perimeter using symbols (e.g. a + a + 3 + 3; 4 a + 3) or assumes a = 5 or 6 cm (e.g., 10 + 3 + 10 + 3 or 12 + 12 + 3 + 3)
2	Correct response but not in its simplest form (e.g., 2a + 2a + 3 + 3)
3	Correct symbolic response in simplified form (e.g., 4a + 6 or 2(2a + 3))

ATILP3

SCORE	DESCRIPTION
0	No response or irrelevant response
1	Incorrect but partially identifiable (e.g., 6a + 6), with little/no working or explanation to support response or incorrect calculated response based on a = 5 or 6 cm
2	Incorrect response due to minor errors but with working that shows understanding of perimeter or correct based on a = 5 or 6
3	Correct symbolic response but not simplified (e.g., a + 1 + 1 + a + 1 + a + 1 + a – 3 + a – 1 + a + 1 + 2 + 1) and without clear explanation
4	Correct symbolic response (6a + 4) without clear explanation (e.g., just added all the sides together) OR correct response but not simplified with a reasonable explanation/working to support solution
5	Correct symbolic response (6a + 4) with clear explanation for sides that are less than a (e.g., a – 1 or a - 3) or explanation based on visualisation.

Figure 1 Three items and their respective scoring rubrics from the Algebra Tiles (ATILP)

The 'big ideas', hypothesised learning progressions, and trial assessment tasks were interrogated by the full research team at an extended face-to-face meeting prior to the first residential professional learning workshop for Specialists and project partners in November 2014. Following the workshop, Specialists were asked to trial the assessment tasks and scoring rubrics and provide advice about their suitability as assessment tasks or teaching activities. This had some unanticipated outcomes. For example, an assessment task that had been used in various studies since 1996 (Batanero et. al., 1996; Watson & Callingham, 2014; 2015) without comment, attracted the ire of one State system. The task presented the results of an actual survey of 250 people in a two-way table to explore the link between smoking and lung cancer. In this case, the inclusion of this task was objected to because the table showed 'no apparent association between smoking and lung cancer' and this was deemed to be misleading. The task was dropped.

While this process provided valuable information about the readability and suitability of the items for the students in the research schools, it did not yield information about the types of reasoning hypothesised to be at the upper end of the progressions. To this end, and to test the suitability of the assessment tasks more broadly, multiple task booklets, referred to as *forms*, were prepared and trialed in a range of non-project schools sourced through AAMT. Each form comprised five to six tasks – either from the one area (Standard Forms) or from two areas (Mixed Forms). Common tasks were included across forms to support analysis as well as anchor items from the SNMY assessment options (Siemon et al., 2006a) to investigate the relationship between mathematical reasoning and multiplicative reasoning. A total of 24 forms were created for this purpose and sent to non-project schools, four for each area of reasoning (e.g. Stats1, Stats2, Stats3, and Stats4) and four for each of the possible combinations of two areas (e.g. Alg/Geo1, Alg/Geo2, Alg/Geo3, and Alg/Geo4).

The trialing generated interest from a diverse range of schools from different States and Territories resulting in over 1000 responses to the various forms from students across Years 5 to 10. The de-identified forms were returned to the research team for marking by trained assessors to test the clarity and sufficiency of the partial credit scoring rubrics. Trial data were analysed using Master's (1982) Rasch partial credit model and Winsteps 3.92.0 (Linacre, 2016). The resulting ordered lists of item rubrics were used to review and refine the hypothetical learning progressions in each area, the refined versions of which were referred to as *Draft Learning Progressions* (DLPs). This process identified some gaps in the progressions that prompted the redesign of some items and/or rubrics and/or the design of additional items to further test and elaborate the draft progressions.

Phase 2

This phase of the RMFII project focused on the design and trial of additional assessment items to test under-evidenced aspects of the DLPs. This led to the preparation of a revised set of mathematical reasoning forms (MR1) that were used by research schools between September 2016 and March 2017. A total of 12 forms (six Standard and six Mixed) with common tasks were created for this purpose (see Table 1).   The forms were marked and moderated by project school teachers using the scoring rubrics provided. The de-identified results were entered into an excel spreadsheet provided by the research team and returned to the team for analysis. Valid responses were obtained from over 3360 students and analysed using the Rasch partial credit model (Masters, 1982) and Winsteps 3.92.0 (Linacre, 2016). Following this, project school teachers were asked to provide written feedback on the suitability or otherwise of the tasks and scoring rubrics, their experience in relation to administering, marking and moderating the MR1 forms , and their thoughts about what the results might mean for them.

Table 1 Indicative sample of tasks and Items by forms to show spread (MR1 RMFII schools)

		Standard Forms						Mixed Forms					
Task	Items	Alg	Alg	Geo	Geo	Stat	Stat	Alg/Geo		Stat/Alg		Geo/Stat	
		A	B	A	B	A	B	A	B	A	B	A	B
Balance	1, 2	o						o		o			
Balloons	1,2,3								o		o		
Rectangle	1,2,3			o	o			o				o	
Symmetry	1,2,3			o				o	o				
Coins	1A,1B						o						o
	2,3					o					o		
Homework	A,B,C					o	o				o	o	

Although the fit of all items to the Rasch partial credit model was acceptable using Linacre's (2019) criteria, some items did not provide sufficient spread of information. For example, an item about designing packaging for soft drinks (GBEV1) had a three-step rubric, the response codes for which all appeared in Zone 7 of the DLP. The item was subsequently omitted as it did not provide sufficient spread at the top of the scale. This prompted a further round of assessment (MR2) in April-June of 2017 to 'flesh-out' certain aspects of the DLPs and trial tasks and scoring rubrics amended as a result of teacher feedback or the Rasch analysis. The data from MR1 and MR2 were then considered together to refine the DLPs and inform the development of the targeted teaching advice. This phase also included a student survey (Barkatsas & Orellana, 2019a, 2019b) and a post MR2 teacher survey to record teacher reflections and identify likely affordances and constraints of using the assessments and targeted teaching advice to inform their teaching.

Phase 3

The final phase of the project collected and analysed data from two further assessment rounds (MR3 and MR4), which were undertaken in late 2017 and early 2018 to determine the extent to which project school teachers could use the draft targeted teaching approach to mathematical reasoning. This phase also focused on the development of professional learning modules to support a targeted teaching approach to mathematical reasoning, and publication of project outcomes and reports for publication on the web-based platform hosted by AAMT.

While the outcomes of this work are described elsewhere (e.g. Siemon, Barkatsas, & Seah, 2019), this paper will focus on a key part of Phase 2, the role of project school teachers in the trial and refinement of the assessment tasks and scoring rubrics and the development of targeted teaching advice from the analysis of student responses to the final assessment forms.

The teachers' role

As indicated above, teachers were encouraged to provide feedback on the assessment items and on how well the rubrics represented the students' mathematical reasoning over the course of the project. The feedback was both in the form of formal surveys as well as anecdotal comments provided with data sent to the research team. The post MR1 teacher feedback survey included seven questions about using the mathematical reasoning forms; the marking moderation process, how teachers were using the results of the mathematical reasoning testing to inform their teaching, how local factors in their school impacted the project, how sustainability of the approach to mathematics teaching could be achieved, the external factors that impacted the project and an open question about sharing exciting or unexpected results. The first three questions, in particular, informed the project as they provided information that influenced the reworking of assessment tasks, rubrics and the targeted teaching advice.

The teachers in this project fulfilled two very important roles, firstly as co-researchers and secondly as reflective practitioners. Their involvement was crucial as Australian teachers have considerable discretion about the mathematics that is taught and how it taught and assessed. Engaging teachers in the enterprise of developing evidenced-based learning progressions, assessment tools, and targeted teaching advice in relation to important aspects of school mathematics, not only ensures the resources are fit for purpose, but it also brings teachers face-to-face with student's reasoning and in so doing becomes a powerful agent for change (Carpenter et al., 2004).

Teachers as co-researchers

The process of task development demanded close cooperation between all stakeholders—researchers, psychometricians, teachers, and education system personnel. As indicated above, there were several iterations of task writing and trialing over the course of the project underlining that this process was far from trivial. Changes to the tasks, rubrics, and the way they were presented were made in response to feedback from trial assessors (phase 1), project school teachers (phase 2), and in some cases by project partners. Research team members, reflecting on the feedback and the results of the Rasch item analyses, also contributed to this process and were responsible for making the changes.

This can be illustrated by a changes to a task designed to focus on reasoning about the relationship between length and volume and between length and surface area. The task, with two items, was first used in the initial trialling phase in February 2016.

a. Mat thinks that if you double the lengths of the edges of a cuboid, its volume doubles. Explain (use a diagram if necessary) if he '**is'** or **'is not' correct.**

b. Mat then says that if you double the lengths of the edges of a cuboid, its surface area doubles. Explain (use diagram if necessary) if he '**is'** or **'is not' correct.**

In both cases the scoring rubric used was 0 – incorrect answer , 1 – correct answer but no or unclear justification, and 2 – correct answer with clear justification.

Following discussions with assessors and teachers the question was placed in the context of a shoebox, the rubrics were altered to include examples to assist teachers (April 2016). The revised rubrics are shown below.

Shoebox Volume

CODE	DESCRIPTION	EXAMPLES
0	Agree or no attempt	
1	Disagree but not full correct reasoning	May say × 4
2	Disagree with sound reasoning	Side lengths doubling mean volume is × 8 (2 × 2 × 2 for doubling each dimension)

Shoebox Surface Area

CODE	DESCRIPTION	EXAMPLES
0	Agree or no attempt	
1	Disagree but not full correct reasoning	May think increase is larger because of the 6 faces
2	Disagree with sound reasoning	Each face will increase area by 4 times so overall ×4 increase

Figure 2 Score codes for the Shoebox Tasks (April 2016)

While the formatting of this was changed for MR1, feedback from teachers after MR1 led to a further clarification of the rubric to ensure the emphasis on reasoning was clear. The revised task items and rubrics are shown in Figure 3.

[GSZLV]

Matt said that if you double the length of the edges of a shoe box, it will double in volume.

Do you agree? Explain your reasoning (You may use diagrams if you wish).

GSZLV

SCORE	DESCRIPTION
0	No response or irrelevant response
1	Agrees with Matt (incorrect) with little/no explanation OR disagrees (correct) with no explanation
2	Agrees with Matt giving reasons that only enlarge one or two dimensions (e.g., may double length only not all edges)
3	Disagrees with Matt with partial reasoning (e.g., may say that it is quadrupled)
4	Disagrees with Matt reasoning based on doubling all side lengths (e.g., *doubling all sides means volume is 8 times larger*), may use formula to show V = 2l x 2 w x 2h

[GSZLSA]

Matt then says that if you double the lengths of the edges of a shoe box, its surface area will double. Do you agree? Explain your answer (You may use diagrams if you wish).

GSZLA

SCORE	DESCRIPTION
0	No response or irrelevant response

1	Agrees with Matt (incorrect) with little/no explanation OR disagrees (correct) with no explanation
2	Agrees with Matt giving reasons that only enlarge one or two dimensions (e.g., may double length only not all edges)
3	Disagrees with Matt with little reasoning (e.g., may think increase is larger because of the 6 faces)
4	Disagrees with Matt, reasoning recognises that each face will increase area by 4 times so overall increase is quadrupled, may use formula to show that surface area quadrupled

Figure 3. Shoebox task and the modified scoring rubric.

After the third round of assessment (MR3) and discussions with teachers at the March 2018 workshop, two further changes were made to what was thought to be the final form of this task. A simple picture of a shoebox was added, and the word “all” was added to GSZLA so that it now reads “Matt then says that if you double the length of *all* the edges of a shoe box, its surface area …”. This change was made on the basis of teacher feedback that “for some students ‘lengths’ prompted them to only consider doubling the length of the box”.

Researchers themselves changed some items on the basis of the content. For example, a task about house prices that had been based on a newspaper article was out-of-date in terms of the price of houses. New information was sought, and the item was rewritten but used the same rubric as before. Another that used data on attitudes to decriminalising marijuana usage was changed to attitudes on the reduction of greenhouse gases using the same numbers and scoring rubric.

Changes to the partial credit scoring rubrics were generated by teacher feedback and by the research team on the basis of the Rasch item fit analysis. This could involve a contraction (e.g. combining a score of 2 and a score of 3) or an expansion as shown in the revised rubrics for the Shoebox problem above. However, at some point the decision about what score is warranted comes down to teacher judgement. A case in point is the student’s response to ATILP2 shown in Figure 4.

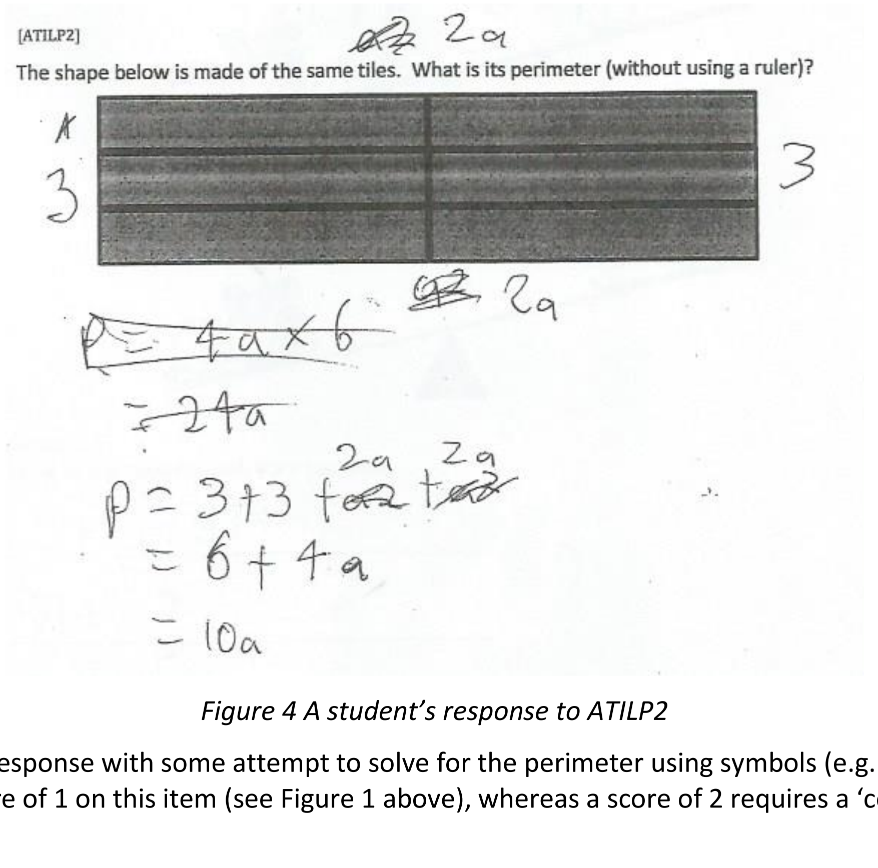


Figure 4 A student’s response to ATILP2

An incorrect response with some attempt to solve for the perimeter using symbols (e.g. a + a + 3 + 3) attracts a score of 1 on this item (see Figure 1 above), whereas a score of 2 requires a ‘correct response but not in its simplest form (e.g. 2a + 2a + 3 + 3)’. Given this, project funds were provided to enable research school teachers to meet together in school time to mark and moderate student responses. Although this was aimed at ensuring the rubrics were applied consistently, this process was invaluable in generating discussions about student reasoning and what was important in school mathematics. For instance, one of the issues that arose in relation to the Algebra Tile problem was the higher score(s) given for simplest form in ATILP3 (see Figure 1 above). Many teachers reported that they felt it was ‘unfair’ to penalise students who had answered the question correctly but had not expressed this in simplest form. This was discussed with research school teachers in an online professional learning session about what was meant by “use as much mathematics as you can”. The issue was resolved by everyone agreeing that a disposition to look for and express relationships in simplest form was a fundamental mathematical practice that supported reasoning with mathematical objects, hence the higher score was warranted.

The feedback from teachers allowed the research team to see the tasks from a teacher’s perspective resulting in some tasks being reclassified as teaching tasks rather than assessment tasks. For example, initially two geometrical reasoning tasks were written on viewing 3D objects from different perspectives. The first was the *Dog’s Eye View* task, which became one of the final assessment task. The second task was *Building Views*. After the first round of trialling these assessment tasks teacher feedback commented on the *Dog’s Eye View* task:

> *Students seemed to grasp page 1 with general ease and were able to understand how to answer the questions*
>
> *This was well received by the students.*

However, for the *Building Views* task teachers commented:

> *Students quickly gave up on and needed ongoing encouragement.*
>
> *Very difficult task that students struggled to get into*

with many saying that the students definitely needed to use concrete materials. On the basis of this feedback, the second task became a classroom activity and the first was retained as an assessment task with minor language changes.

In other cases the modification to a task was small but had a significant impact on student understanding. For example, in the Boxes task that involved working out which of the boxes was heavier than the others using reasoning with a balance, the feedback was:

> *Task is easy to use, can be used as an assessment for reasoning. In general though the students did not understand that the numbers on the boxes did not indicate the weight or reflect the weight, so we suggest a sentence is added at the beginning to state that the numbers do not reflect the weight.*

This was resolved by labelling the boxes with letters rather than numbers.

Another example was provided by a research school teacher in response to the R*elations* task: “ARELS1 – [I] thought there was a descriptor missing from the rubric – 2 [marks] for noting the difference of 6, but 3 [marks] for noting which number had to be six more (LHS or RHS)”. Figure 5 shows the ARELS1 task and the rubric that was modified in response to this feedback.

[ARELS1]

What numbers would go in these boxes to make a true number sentence (the numbers may be different). Explain your reasoning.

$\square + 521 = 527 + \square$

ARELS1

SCORE	DESCRIPTION
0	No response or irrelevant response
1	Incorrect response but suggest the difference of 6 is recognised in some way (e.g., *add 6 to the right hand side*)
2	Two correct numbers given (e.g., 13 and 7; 527 and 521) but little/no reasoning.
3	Two correct numbers given where the number on the left is 6 more than the number on the right (e.g., 100 and 94) with reasoning that reflects the relationship between 521 and 527 (difference of 6).

Figure 5. ARELS1 task and the modified scoring rubric.

The close working relationship among the researchers, psychometricians, teachers, and industry partners allowed a comprehensive set of tasks to be developed that resonated with practitioners and had sound psychometric properties (Siemon & Callingham, 2019). Teachers commented that students engaged well with the assessment tasks, taking them seriously and working to their best ability. Additionally, assessment tasks deemed as unsuitable for assessment on various grounds, provided quality classroom activities that reinforced the desired focus on reasoning.

Teachers as reflective practitioners

Teachers commented that marking their own students' work provided them with insights that they would not otherwise have had. For example, in response to a question on the post MR1 survey, two teachers provided the following comments.

> *Algebra for year 9s is about to be taught now. Whilst the curriculum is about factorising and expanding brackets, I feel these results indicate that it will be important for the team to do some pattern recognition work, such as Max's match sticks (SNMY task). The team will be meeting next week to discuss what will be key for the students to cover now and then to leave and ensure they cover next year as they head into year 10. It will be interesting to see what the results are for the groups for the forms in a few weeks after this short topic.*
>
> *The testing highlighted some misconceptions that the students had and that the teachers believed they didn't have. I tested my students after the statistics unit and realised that some concepts hadn't been embedded as I had thought and so I went back and taught them through other methods to try to eliminate this problem – it would be interesting to test the kids again.*

Many schools chose to set aside staff meeting time so that teachers involved could work collaboratively on the marking. This process provided an opportunity to unpack the rubrics and to moderate the scoring. This was seen to be a very important part of the process that had benefits beyond simply marking the work. One teacher wrote:

> *The moderation was good for our staff and we quickly reached consensus. The exception was one staff member who had difficulty relating the student's answers to the rubric. The staff member would look at the work and judge it by criteria that didn't relate to the rubric. "This student is trying really had and I can see they are quite clever; they should get a higher mark". Initially we attempted to instruct the teacher towards the model we were using however we eventually realised there was a pedagogical difference that needed addressing.*

Another teacher wrote: "There were moments of awareness and realisation during the moderation process … It was a really valuable process to make staff aware of gaps that are current across the whole range of students."

The teachers demonstrated a willingness to engage with the notion of learning progressions and the ways in which they could be used to better target their teaching. Once the draft Learning

Progressions had been established following the implementation of assessments MR1 and MR2, they were presented to teachers by the psychometrician at the residential workshop at the end of 2016. At this meeting the process of analysis was explained, and teachers shown the Wright (variable) maps produced by the software. The ways in which rubrics mapped onto the variable, and how these were then used to develop the zones and zone descriptors were elaborated. Challenges arising from the data were also explored. For example, the distribution of students along the Geometric Reasoning and Statistical Reasoning variables indicated that students found these two areas much more challenging than Algebraic Reasoning. Few students appeared in Zone 7 and Zone 8 of these variables, and in Geometric Reasoning there were also few items in the lower zones. These findings suggested that more easy geometry items were needed to ensure there was sufficient information about the whole variable. Teachers recognised that more work was needed in statistics and geometry. They also commented that the curriculum stressed algebra over the other domains and that they sometimes ran out of time to address statistics and geometry.

Using work samples from their students that they had brought to the workshop, teachers engaged in an activity to create a mini learning progression in collaboration with teachers from a different school. By sharing high, middle and low-level responses, they developed a description of student characteristics at these levels in one of the domains. In this way teachers developed a deeper understanding of the psychometric process, and this reduced the sense of the assessment development and subsequent learning progression being removed from the classroom.

In feedback obtained from teachers they also commented on particular questions. For example, an item called *Movie World* had a focus on sampling. Students sometimes missed important contextual information in the question that had an impact on the sample. Teachers recognised that this was happening but commented that recognising important contextual information is a crucial element of Statistical Reasoning. One teacher indicated that one way of addressing this issue was to develop and run a real survey in their school.

The development of targeted teaching advice

Once there was a stable construct for each of the three reasoning scales and all of the data were analysed using Rasch analysis (Bond & Fox, 2015) the research team met to interrogate the data to identify what student behaviours were evident in each of the zones of the learning progressions. During this process student behaviours that needed to be consolidated and established in each zone were identified. Then concepts that needed to be introduced and developed in order to move the students to the next zone were identified. It was evident from these discussions that there was significant overlap between zones and that what was introduced and developed in one zone was what could be consolidated and established in the next zone in the progression.

Broad descriptions for each zone that included what students in a zone should be able to do and what they may find challenging were developed. Each of the three Learning Progressions are based around three main Big Ideas. The broad descriptions were then fleshed out to provide teachers with enough information to use the targeted teaching advice but not so much as to overwhelm them. For example, in Statistical Reasoning the targeted teaching advice is arranged around the three Big Ideas of:

- Variation with Expectation and Randomness (VER)
- Variation with Distribution and Expectation (VDE)
- Variation with Informal Inference (VII)

As an example, the targeted teaching advice for Zone 6 in Statistical Reasoning is shown in Table 2.

Table 2*Targeted Teaching Advice for Statistical Reasoning Zone 6*

ZONE 6 Description	**Teaching Implications**

VER *Constructs reasonable arguments based on an understanding of chance and probability (SCON3.3) and context (SBED.2).* *Uses measures of central tendency to justify a closed response (AMED.3; SHSE3.1).* ***VDE*** *Interprets and describes the association between two variables and considers the implication (e.g., SHWKB.3; SM8GR.2) in visual contexts.* *Beginning to work with the association between two variables in non-graphical format (SSKIN.3).* ***VII*** *Provides sensible critique of sampling in context of method and sample size but is implicit rather than explicit about randomness (SMVE11.2).*	**Consolidate and Establish** ➢ **Critique different graphical representations** and represent relationships in non-graphical forms e.g., 2-way tables, nested tables; Explore the ABS census website and critique different representations of data (http://www.abs.gov.au). ➢ **Represent and quantify relationships between two variables** to make variability explicit, e.g., *Digging into Australian Data with TinkerPlots* datasets. Explore probability distributions using simple experiments or simulations (e.g., racetrack, sum of two dice) to show non-uniform distributions. ➢ Draw pictures of distributions that recognise unevenness/variations in distributions e.g., sketch what you think the distribution of handspans would look like. Collect data and compare with sketches. ➢ **Explore the relationship between statistical data and algebra** e.g., the relationship between handspan and foot length; distance-time graphs. **Introduce and Develop** ➢ Introduce the differences between **observed and expected values** e.g., compare the distribution of heads from 10 coin flips with the theoretical value; *Two Coins [provide link].* ➢ **Simulate real situations** using random generators, e.g., Birth Month Problem (maths300); Coke vs Pepsi taste test https://serc.carleton.edu/sp/cause/datasim/examples/cokepepsi.html. Simulate the test using coin tosses. ➢ **Focus on proportion**, e.g., using % when comparing attributes; visit the ABS Census site and use the data provided to move from % to counts and write a news story about a particular place. *Two coins [provide link].*

After attending both face-to-face and online professional learning run by the RMFII project team, the feedback from project schools was that they wanted to modify the existing pedagogical approach in their schools to provide more effective and engaging ways of teaching their students. In particular, teachers wanted to be able to approach mathematical reasoning across a range of zones with mixed ability groupings of students. Given the research indicating the efficacy of working with groups in flexible ways rather than ability grouping (Boaler et al., 2000; Ireson, & Hallam, 1999; Sullivan, 2011) and the use of low threshold high ceiling tasks (Boaler, 2016; McClure, 2011; Sullivan, 2011) it was decided to link the teaching advice to indicative tasks that teachers could use with students across several zones. Tasks from well-known sources such as https://maths300.com/, https://resolve.edu.au and https://nrich.maths.org/ were linked to the targeted teaching advice.

The development of the teaching advice from the learning progressions is included here to demonstrate that they occupy a particular space between the *Australian Curriculum: Mathematics* (AC:M) and teachers' classroom practice. The deliberate choice to focus on 'big ideas' and mathematical reasoning was made to ensure that the progressions would be of a grain size and form that was meaningful to teachers, while at the same time providing a basis for ongoing professional learning.

Discussion

Towards the end of the RMFII project, the specialist teachers from each school and their principals were asked to respond to a survey about how the involvement in the project had affected what mathematics and how mathematics was being taught in their schools. Teachers were quite reflective about the changes they noticed in the classroom climate as a result of implementing a targeted teaching approach. Several of the comments from teachers focused on students, especially the students who were normally disengaged, being willing to try in mathematics classes. Comments such as "The program shaped my classroom environment and has shifted my students' thoughts about maths and feelings about themselves as maths students, from 'bad at maths and won't bother' to 'I can give it a go'." were typical. Several of the comments on the improvement in classroom climate referred to a renewed enthusiasm from teachers as student engagement improved.

Carpenter et al. (2004) proposed that if teachers, and students, are to learn with understanding that they need to make knowledge their own and this is often done when teachers work together. Teachers from the project schools who engaged in the marking and moderation process found that they learned much from working together to generate new knowledge about the mathematics learning in their schools. For example, new units of work were designed as a result of the process:

> *After pre-testing of the Algebra MR1 booklet we as a grade team decided to teach an algebraic reasoning unit to develop students' skills in this area. Upon reflection of our existing unit plan we identified that the learning opportunities for students allowed mostly for fluency (not reasoning/problem solving). The results from the MR assessment were used to direct teaching. We distributed tasks that were aimed at the level of students in relation to their MR booklet result (project* teacher survey response).

To teach for understanding requires teachers to understand how they can scaffold learning to support students to make sense of their mathematics (Carpenter et al., 2004). One of the project schools identified a specific aspect to work on as a result of the moderation process. "We have tried to address the … more variation in smaller samples with Year 9s, using data collected from tossing dice, comparing individual results with the combined class results. We would not have recognised the need for this without the testing."

As well as addressing particular ways of teaching content, teachers were reflective about their pedagogical content knowledge (PCK) (Shulman, 1986) by engaging in reflective dialogue and sharing their teaching practices with others (Carpenter et al., 2004). When asked what the outcomes of the school's participation in the project for the teacher, one teacher responded: "Collaborative work with another teacher; recognising how to move students on from one level to the next; how to use ordinary equipment in other ways that enhance student learning (e.g. rope and pegs); and student engagement gives great joy, and less behaviour management issues." A principal in another school stated that an outcome for the teachers in their school was "An improvement in pedagogy that creates rich learning environments that engage all students in mathematics." Several of the comments from the teachers involved in the project showed that their foci had changed, and they were moving away from more traditional approaches to teaching mathematics to using rich tasks, with multiple entry and exit points, at a level that was challenging for individual students, to address gaps in student learning.

The principals involved in the project recognised that both the teachers and the students were exhibiting increased capacity to learn about mathematics. Comments such as "Building the capacity of teachers' pedagogy to better engage students and build their maths knowledge, understanding, skill acquisition, problem solving ability and to improve teachers' ability to differentiate the learning" were typical of the principals' comments about their teachers' involvement with the project. When asked to reflect on the impact of student engagement a typical comment was "They enjoy classes more, understand the application of math better and have been scaffolded in their learning more." One principal even suggested that student attendance was better on the days set aside for project work. Several principals also commented on how teachers focused collectively on student learning and how it was this sense of a shared purpose and collaboration that was affecting the improvement on teaching and learning (Carpenter et al., 2004) that was being seen in the schools.

One principal stated "[In the future] we will continue as we are and using this year's participants to grow the knowledge and understandings (capacity) of other staff and reinforce that it is everyone's responsibility to be a teacher of both Literacy and Numeracy." This speaks to the importance of leadership in sustaining genuine reform (Carpenter et al., 2004; Leith wood, et al., 2004).

The way in which the teachers were centrally involved in contributing to the development of assessment tasks, rubrics and teaching activities in this study is a point of difference from other major learning progression work in mathematics education. For example, in the *Learning through an Early Algebra Progression* (LEAP) Project the research project team designed the tasks and used external scholars as reviewers and then a member of the research team taught the curriculum sequence (Blanton et al., 2019; Fonger et al., 2018).

A number of authors have sought to describe the nature of the relationships between researchers, psychometricians, and practitioners involved in complex enterprises such as the one reported here. Notions of boundary crossings boundary practices, and boundary objects have been around for many years but more recently, these have been presented as a useful way of thinking about the relationship between learning progressions, learning, and assessment (Lehrer et al, 2014), and for conceptualising research-practice partnerships like those involved in the RMFII project as "joint work at boundaries" (p. 182). While there is more work to be done to theorise the nature of the relationships involved in the RMFII project and ongoing relationships with the professional community as these outcomes are shared, this would appear to be a useful s and the role of artefacts such as the assessment forms in providing objects around which different agents, teachers, researchers, students, can come together to explore what is involved in mathematical reasoning at the classroom level, this looks like a promising way forward.

Conclusion

The RMFII project set out to involve practitioners in the development of evidence-based, formative assessment materials for mathematical reasoning that could be translated to practice at scale (Cobb & Jackson, 2011). One of the motivations for doing this was to support a sustained change in practice away from low-complexity, procedural exercises to teaching based on a deeper understanding of the big idea and the connections between them (Sullivan, 2011). By partnering with the AAMT and State and Territory education authorities, the materials are in the public domain, but it is the processes around these that teachers, school leaders and policy makers engage in that will determine the extent to which these impact teacher's everyday decision making and students mathematics learning. Coming to know and understand the big ideas that underpin each of the three reasoning areas can help teachers make connections to other areas of mathematics. They also provide an organising frame to help teachers make decisions about what mathematics is important and how best to spend their time and the time of their students in mathematics classes. And, in bringing teachers together in collaborative groups to mark and moderate, the assessment forms bring teachers face-to face with evidence of students learning and reasoning to provide a strong basis for school-based professional learning.

All of which raises the issue of dissemination and scaling up. Publishing papers and book chapters is unlikely to bring about change in the short term, although these are important. Wiliam (2006) emphasised the critical importance of professional learning in sustaining an evidence-based approach to teaching and learning mathematics. But what does this mean in practice? Placing "student reasoning at the centre of instructional decision making" (Carpenter et al, 2004, p. 14) is known to be what 'travels' - the RMFII resources provide the means by which to do this, but how to connect teachers to the resource remains an issue. The partnership with AAMT means that the resources will be available for free on the AAMT website. While this would seem to offer an effective means of dissemination, AAMT has by its nature many other resources on its websites, and like all websites there is the issue of attracting people to the website. Included in the suite of resources are some professional learning modules, the aim of which is to support school-based, teacher learning communities explore and use the resources to make better, more informed decisions about what to teach and how they might teach it to more fully engage students in the enterprise of learning mathematics. Perhaps from small things …

> *It's too early to tell yet, but I think we are slowly but surely building cohorts of students that are moving from superficial understandings to stronger, more grounded and deeper understandings and love of Mathematics…at least that is my dream, and I won't give up!* (MR1 teacher feedback)

References

Australian Curriculum Assessment & Reporting Authority [ACARA]. (2016). Australian Curriculum: Mathematics F-10. https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics/

Australian Industry Group. (2015). *Progressing STEM skills in Australia.* Melbourne: AIG

Australian Mathematical Sciences Institute [AMSI]. (2020). *The state of mathematical sciences 2020*. https://amsi.org.au/wp-content/uploads/2020/05/amsi-discipline-profile-2020.pdf

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of Learning Sciences, 13*(1), 1-14.

Barkatsas, T., & Orellana, C. (2019a). Investigating mathematics students' motivations and perceptions. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 205-220). Brill Sense. https://doi.org/10.1163/9789004396449_010

Barkatsas, T. & Orellana, C. (2019b). Secondary students' mathematics education goal orientations. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 221-231). Brill Sense. https://doi.org/10.1163/9789004396449_011

Baroody, A., Cibulskis, M., Lai, M., & Li, X. (2004). Comments on the issue of learning trajectories in curriculum development and research. *Mathematics Teaching and Learning, 6*(2), 227-260.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*, 151-169.

Black, P., Wilson, M., & Yao, S. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives, 9,* 71-123.

Blanton, M., Isler-Baykal, I., Stroud, R., Stephens, A., Knuth, E., & Murphy Gardiner, A. (2019). Growth in children's understanding of generalizing and representing mathematical structure and relationships. *Educational Studies in Mathematics*, *102*, 193–219.

Boaler, J. (2006). How a detracked mathematics approach promoted respect, responsibility, and high achievement. *Theory into Practice, 45*(1), 40-46.

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teachers.* Jossey-Bass.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping: Disaffection, polarisation and the construction of failure. *British Educational Research Journal, 26*(5), 631-649.

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Lawrence Erlbaum.

Business Council of Australia. (2015). *STEM strategy would help drive innovation*. [Media Release], September: http://www.bca.com.au/media/stem-strategy-would-help-drive-innovation

Carpenter, T. P., Blanton, M. L., Cobb, P., Franks, M. L., Kaput, J., & McClain, K. (2004). *Scaling up innovative practices in mathematics and science*. National Center for Improving Student Learning and Achievement in Mathematics and Science.

Clements, D. H. (2007). Curriculum research: Toward a framework for research-based curricula. *Journal for Research in Mathematics Education, 38*, 35-70.

Clements, D. H. & Samara, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81-89.

Cobb, P., Confrey, J., di Sessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13. http://**doi**.org/**10.3102/0013189X032001009**

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education Development, 13*(1), 6-33.

Confrey, J. & Maloney, A. P. (2010). The construction, refinement and early validation of the equipartitioning learning trajectory. In K. Gomez, L. Lyons & J. Radinsky (Eds.), *Proceedings of the Ninth International Conference on the Learning Sciences*, Vol 1 (pp. 968-975). Chicago, IL: ICLS.

Daro, P., Mosher, F. A., & Corcoran, T. B. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment and instruction.* (Research Report No. 68). Consortium for Policy Research in Education. https://files.eric.ed.gov/fulltext/ED519792.pdf

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

Fonger, N., Stephens, A., Blanton, M., Isler, I., Knuth, E., & Murphy Gardiner, A. (2018). Developing a learning progression for curriculum, instruction, and student learning: An example from mathematics education. *Cognition and Instruction, 36*(1), 30-55.

Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education, 25*(3), 343-358.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children to learn mathematics.* National Academy Press.

Lehrer, R., Kim, M., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In A. Maloney, J. Confrey, & K. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 31-59). Charlotte: Information Age Publishing.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning. Review of research.* The Wallace Foundation.

Linacre, J. M. (2016). *Winsteps Rasch Measurement* (Version 3.92.0) [Computer software]. Rasch.

Linacre, J. M. (2019). *Winsteps online help manual*. Winsteps.

Masters, G. (1982). A Rasch model for partial credit modelling. *Psychometrika, 47,* 149-174.

Masters, G. (2013). Reforming educational assessment: Imperatives, principles and challenges. *Australian Education Review, 57*, 1-57.

McClure, L. (2011). *Using low threshold high ceiling tasks in ordinary classrooms.* https://nrich.maths.org/content/id/7701/LTHCArticle.pdf

Office of the Chief Scientist. (2016). *Australia's STEM workforce: Science, technology, engineering and mathematics*. Australian Government. https://www.chiefscientist.gov.au/sites/default/files/Australias-STEM-workforce_full-report.pdf

Office of the Chief Scientist, (2017). Science and maths in Australian Secondary Schools – Datasheet 2. https://www.chiefscientist.gov.au/sites/default/files/2-Science-and-Maths-in-Australian-Secondary-Schools-datasheet-Web.pdf

Penuel, W. R., Allen, A., Coburn, C. E., & Farrell, C. (2015). Conceptualizing research-practice partnerships as joint work at boundaries. *Journal for Education for Students Placed at Risk, 20*(1-2), 182-197. http://dx.doi.org/10.1080/10824669.2014.988334

PricewaterhouseCoopers. (2015). *A smart move: Future proofing Australia's workforce by growing skills in science, mathematics, engineering and maths (STEM).* http://www.pwc.com/gx/en/industries/government-public-services/public-sector-research-centre/australia/smart-move.html

Shepard, L. A., Penual, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice, 37*(1), 21-34.

Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics, 82*(2), 21-34.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Siemon, D. (2013). Launching mathematical futures: The key role of multiplicative thinking. In S. Herbert, J. Tillyer, & T. Spencer (Eds.) Mathematics: Launching futures. *Proceedings of the 24th biennial conference of the Australian Association of Mathematics Teachers* (pp. 36-52). Melbourne: AAMT.

Siemon, D. (2016). *Addressing the STEM challenge through targeted teaching – What's the evidence.* Invited presentation to the annual Research Conference of the Australian Council of Educational Research, Brisbane (August), http://research.acer.edu.au/cgi/viewcontent.cgi?article=1277&context=research_conference

Siemon, D. (2017). Developing learning progressions to support mathematical reasoning in the middle years: Introducing the Reframing Mathematical Futures II project. In A. Downton, S. Livy, & J. Hall (Eds.), 40 years on: We are still learning. *Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia* (pp. 651-654). Melbourne: MERGA.

Siemon, D. (2019). Knowing and building on what students know: The case of multiplicative thinking. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 6-31). Brill Sense. https://doi.org/10.1163/9789004396449_002.

Siemon, D. Banks, N., & Prasad, S. (2018). Multiplicative thinking a STEM foundation. In T. Barkatsas, N. Carr & G. Cooper (Eds.), *STEM education: An emerging field of inquiry*. Sense Publications.

Siemon, D., Barkatsas, T., & Seah, R. (Eds.) (2019). *Researching and using learning progressions (trajectories) in mathematics education.* Brill/Sense.

Siemon, D., Bleckly, J., & Neal, D. (2012). Working with the big ideas in number and the Australian curriculum mathematics. In W. Atweh, M. Goos, R. Jorgensen, & D. Siemon (Eds.), *Engaging the Australian curriculum mathematics – Perspectives from the field* (pp. 19-46), Mathematics Education Research Group of Australasia.

Siemon, D., Breed, M., Dole, S., Izard, J., & Virgona J. (2006a). *Scaffolding Numeracy in the Middle Years – Project findings, materials and resources, Final report*. Retrieved from http://www.education.vic.gov.au/school/teachers/teachingresources/discipline/maths/assessment/Pages/scaffoldnum.aspx

Siemon, D., Breed, M., Izard, J. & Virgona, J. (2006b). The derivation of a learning and assessment framework for multiplicative thinking, In J. Novotna, H. Moraova, M. Kratka & N. Stehlikova (Eds.), *Proceedings of the 30th conference of the International Group for the Psychology of Mathematics Education* (Vol.5, pp.113-120). Prague: PME.

Siemon, D. & Callingham, R. (2019). Researching mathematical reasoning: Building evidence-based resources to support targeted teaching in the middle years. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 101-125). Brill-Sense. https://doi.org/10.1163/9789004396449_006

Siemon, D., Callingham, R., Day, L., Horne, M., Seah, R., Stephens, M., & Watson, J. (2018). From research to practice: The case for mathematical reasoning. In J. Hunter, P. Perger, & L. Darragh (Eds.). Making waves, opening spaces. *Proceedings of the 41st annual conference of the Mathematics Education Research Group of Australasia* (pp. 40-49). Auckland: MERGA.

Siemon, D., Day, L., Stephens, M., Horne, M., Callingham, R., Watson, J., & Seah, R. (2017). Reframing Mathematical Futures: Using learning progressions to support mathematical thinking in the middle years. In A. Downton, S. Livy, & J. Hall (Eds.), 40 years on: We are still learning. *Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia* (pp. 646-662). Melbourne: MERGA.

Siemon, D., & Virgona, J. (2001). Roadmaps to numeracy – Reflections on the Middle Years Numeracy Research project. *Proceedings of the annual conference of the Australian Association for Research in Education.* Fremantle: ACER. https://www.aare.edu.au/data/publications/2001/sie01654.pdf

Simon, M. (2006). Key developmental understandings in mathematics: A direction for investigating and establishing learning goals. *Mathematical Thinking and Learning, 8*(4), 359-371.

Stacey, K. 2010). Mathematics teaching and learning to reach beyond the basic. In ACER (Ed.), *Proceedings of 2010 ACER research conference* (pp. 17-20). Melbourne: ACER.

Sullivan, P. (2011). *Teaching mathematics using research-informed strategies*. Australian Council for Educational Research.

Sullivan, P. (2012). The Australian curriculum: Mathematics as an opportunity to support teachers and improve student learning. In B. Atweh, M. Goos, R. Jorgensen & D. Siemon, (Eds.). *Engaging the Australian curriculum mathematics – Perspectives from the field* (pp. 175-189). Mathematics Education Research Group of Australasia.

Swan, M., & Burkhardt, H. (2012). A designer speaks – Designing assessment of performance in mathematics. *Educational Designer, 2*(5), https://www.educationaldesigner.org/ed/volume2/issue5/article19/

Thomson, S., De Bortoli, L., & Underwood, C. (2016). *PISA 2015: A first look at Australia's results.* ACER Press.

Thomson, S., Wernert, N., O'Grady, E., & Rodrigues, S. (2016). *TIMSS 2015: A first look at Australia's results.* ACER Press.

Vincent, J., & Stacey, K. (2008). Do mathematics textbooks cultivate shallow teaching? Applying the TIMSS Video Study criteria to Australian eighth-grade mathematics textbooks. *Mathematics Education Research Journal, 20*(1), 82-107.

Wai, J., Lubinski, D., & Benbow, C. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835.

Watson, J & Callingham, R., (2014). Two-way tables: Issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning, 16*(4) 254-284. https://www.tandfonline.com/doi/full/10.1080/10986065.2014.953019

Watson, J. & Callingham, R. (2015). Lung disease, indigestion, and two-way tables. *Investigations in Mathematics Learning 8*(2), 1-16.

Weldon, P. (2016). *Out-of-field teaching in Australian secondary schools.* ACER Press.

Wiliam, D. (2006). Assessment for learning – what, why and how. *Edited transcript of a presentation to the Cambridge Assessment Network Conference.* University of Cambridge. http://dylanwiliam.org/Dylan_Wiliam_website/Presentations.html

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle, 21*(2), 15-20.