Bringing Research on Students' Understanding into the Classroom through Formative Assessment

Kaye Stacey
University of Melbourne
<k.stacey@unimelb.edu.au>

'Specific Mathematics Assessments that Reveal Thinking' (abbreviated as 'smart tests') provide on-line formative assessment of middle years students. They aim to put information from research on students' understanding directly into the hands of teachers, by providing quick automated diagnosis of learning for all students in a class. The Reflections test is used as an example to describe item presentation, evidence identification, and reporting to teachers, and highlight how pedagogical content knowledge can be built.

Taking research into practice

This paper describes one attempt to put the results of research in education into the hands of teachers. Mathematics education is indeed an interesting subject in its own right, but the primary reason for it to be supported and researched is to help more students learn more mathematics in a deeper, more fulfilling and more useful way. There are many ways in which research can influence practice. Some research findings can directly influence educational policy, at the level of systems or schools. Examples include research on the advantages and disadvantages of streaming or setting by ability, or special provisions for girls' education. These can be taken into practice at the school or system level. Other research contributes to the quality of lessons, of curriculum as experienced by students, and of teaching and learning. This is harder to take into practice, because it affects a myriad of small actions by all teachers every day. A major strand of this research has been to study students' thinking and learning related to specific topics. Beginning in the early 1900's with the study of learning number facts and carrying out arithmetic algorithms, in mid-century it took the 'cognitive turn' to study students' conceptual development, and we now have accumulated a rich understanding of the ways in which students develop their knowledge, skills and understanding of the main topics of school mathematics, and of the features which need special attention in teaching. There are still gaps and important continuing research – about 20% of the 2012 MERGA conference papers demonstrated continuing work in this area – but there is a wealth of information which could improve learning that is locked away in books and not being used by many teachers.

Does it make a difference to student outcomes if teachers' understand how their students think about mathematics? Early research reached the conclusion that this had only a weak effect. However, these studies generally used proxy variables, such as qualifications and course (subject) attendance. In contrast, strong effects have been found in recent studies which have measured the content knowledge of mathematics which is directly involved in teaching and the associated pedagogical content knowledge. For example, Baumert et al (2010) demonstrated a substantial influence of teachers' pedagogical content knowledge (PCK) on students' learning gains over one year in 194 German Year 10 classes. They also found that strong content knowledge (beyond the year 10 content) matters most by enabling the development of good PCK.

In V. Steinle, L. Ball & C. Bardini (Eds.), Mathematics education: Yesterday, today and tomorrow (Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia). Melbourne, VIC: MERGA.

There are many excellent programs which bring research on student understanding to teachers, principally through teacher education, professional development and behind-the-scenes influence on the school curriculum and textbooks. For example, the Australian early years numeracy programs transposed fundamental research on learning into practical programs that increased teachers' understanding of the stages of learning and showed how to use this knowledge in school. This paper describes a different approach to put research into teachers' hands, using new technology at the point of need. It is experimental and incomplete, yet demonstrates what is possible.

Specific Mathematics Assessments that Reveal Thinking

Along with Beth Price, Eugene Gvozdenko and Vicki Steinle (and earlier Helen Chick), I have designed a computer-based assessment tool for teachers of students in approximately Years 5 – 9. 'Smart tests' (an abbreviation of "specific mathematics assessments that reveal thinking") are now being used by many teachers through the website www.smartvic.com. As of June 2013, there are tests on about 60 very specific topics, nearly all with pre- and post-test pairs. The tests are short, completed online, and results are immediately available to teachers. The intention is that teachers will use them just before teaching a topic to better understand the needs of their own students, so they can have a direct influence on the instruction provided. This is one type of formative assessment. Interventions which provide 'assessment for learning' have been shown to have a strong impact on learning outcomes. Some teachers use smart test results to identify groups for specific assistance, and other teachers adapt their plans so that their lessons match more closely where the class is. Perhaps surprisingly, often teachers have reported starting at a more advanced point than they expected (Steinle & Stacey, 2012).

As the diagram in Figure 1 shows, smart tests are intended to impact directly on learning, as teachers better adjust instruction to what the students know, but they are also intended to impact on learning indirectly. As teachers use the tests and act on the diagnoses of students' understanding, it is hoped that they will come to understand in a very practical way how students are likely to think about the topic, the common errors and misconceptions that they might have and how the topic develops from simple to advanced. It is hoped that the outcome will be an increase in teachers' PCK. In this way, the smart tests may become redundant, as teachers take care to develop strong concepts in students, modify their teaching to reduce the likelihood of misconceptions, and have ready access to items which reveal understanding to monitor their classroom practice and student progress.

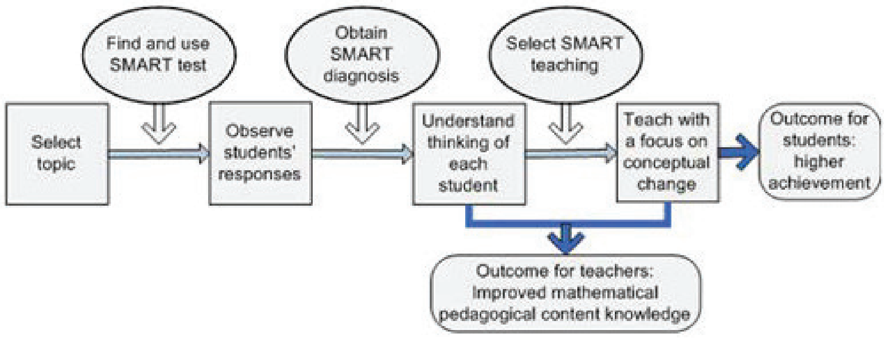

Figure 1. Smart tests are hypothesized to improve learning through two pathways.

An Example: Reflections Smart Test

In this section, many aspects of smart tests are illustrated using the example of a test on Reflections. Other publications have described other aspects of the test design with other sample topics, such as understanding of algebraic letters (Steinle, Gvozdenko, Price, Stacey, & Pierce, 2009), and line graphs (Stacey, Price & Steinle, 2012). There are currently two geometric transformations tests: Reflections and Rotations.

The Australian Curriculum (ACARA, n.d.) describes work on reflections in the 'Location and Transformation' sub-strand. At Year 2, students can explain the effect of a flip (reflection) knowing, for example, that objects move without a change in size or features other than orientation. In Years 3 and 4, students identify line symmetry (e.g. in art and the environment), and make use transformations to make patterns from shapes. In Years 5 and 6, students identify reflections concretely by flipping and folding two dimensional shapes, describe the effect of combinations of transformations, and describe simple reflections using coordinates. By Year 7 they work with generalizations and combinations of transformations, perhaps finding out that the effect of reflection in two parallel mirror lines is a translation. After Year 7, knowledge of transformations is used to test for congruence, and the coordinate work leads to transformations of graphs. The description above establishes that at the end of the middle years of schooling, students should be familiar with three somewhat different aspects of reflections (i) as flipping where the action of turning over is the dominant idea and (ii) as folding or reflecting in a mirror (line of symmetry), where the dominant idea is of two parts, identical except for orientation, and (iii) formalizing this to the mathematical description of an object, a line of symmetry and an image. The Reflections smart test examines the mirror/fold aspect of single reflection transformations.

Figure 2a shows one Reflections items where students drag a tiny 'card' to show the image of a point under reflection. The instruction is:

> "Blobs of paint are dropped on paper. The paper is folded along the black lines. The wet paint makes new copies of the blobs on the paper. Drag [the] blob from the bottom to the correct spot to show how the new copies of the blobs will look." (www.smartvic.com)

The screens are illustrated (with labels added) in Figure 2b. Students can experiment by moving the image blob around the screen, until satisfied with its position. Other items use differently placed blobs and complex shapes, with differently oriented fold lines. As far as possible, smart tests avoid technical language and use tasks that are as realistic as possible. The folding scenario seemed to fit these criteria.

Figure 2. The task of reflecting a point in a vertical line, with dragging illustrated.

The reflections test demonstrates how computer-based assessment can now provide a range of engaging tasks, well beyond the text-based, multiple choice restrictions of some years ago. In this case, drag and drop in a web-delivered test (programmed for desktops and tablets) delivers practical tasks to demonstrate understanding. Even more usefully, the smart test is programmed to assess students' work and identify their common errors, both of which are time consuming work for teachers. The report is immediately available.

By the end of 2012, the Reflections test had been used by 504 students of volunteer teachers. Ten teachers provided most of the data. 88% of the sample is in Year 7 and 12% in Year 8. Figure 3 shows scattergrams of where the images of the blobs were placed by a sub-sample of 90 students. The scattergrams do not show the number of placements at each position, but the spread and density of dots gives a general impression of the responses. Figure 3a shows a solid blob to reflect, the correct placement of the image, and student responses. It demonstrates that reflection in a vertical line is usually well done, although many students do not attend carefully to the distance. Isolated dots probably come from students who did not understand the task, or who did not finish their response. When generous tolerances for judging equal distances were allowed, 82% of students were correct. Figure 3b shows the difference when the fold line is not vertical, with success rate dropping to 25%. Most know that the image will be in the bottom right hand corner, but not exactly where. Many of the images are on the same horizontal level as the original, because many students act as if all folds are vertical or horizontal.

The cluster of placements at the top of Fig. 3b scattergram reveals a different common error. Some students fairly consistently place the image to achieve a pleasing visual balance (in fact it is rotational symmetry about the midpoint of the fold line, but students would not analyse it this way). The small cluster in the lower right hand corner of Fig 3a exhibits the same thinking. It has been reported in the research literature for many years that these are common errors (see, for example, Kűchemann, 1981; Schultz, 1978).

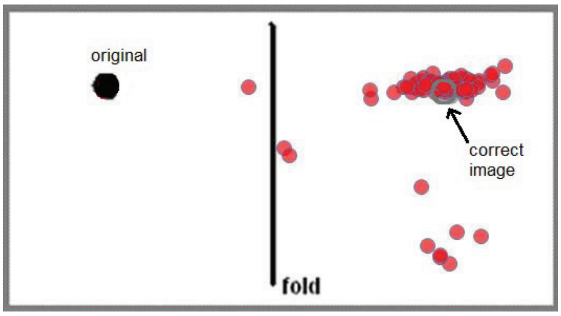

Fig 3a. Placement of images of black circle.

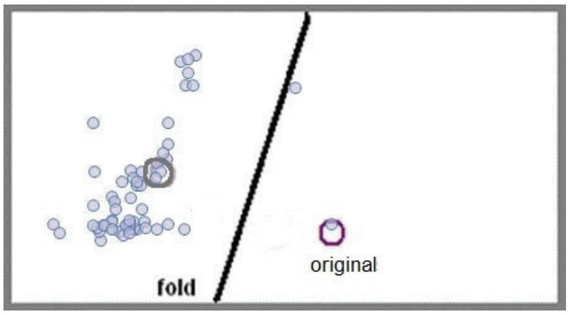

Fig 3b. Placement of images of open circle.

Figure 3. Scattergrams of students' placement of reflections of simple shapes.

Figure 4a shows one of several tasks requiring reflecting a complex shape. Only a handful of students were correct on this item. Students select the fish image that is in the correct orientation and drag it into position. First, they can experiment with the images, moving them anywhere and replacing as required. Currently, rotation of the drag image is not possible in our software. The results confirm the previous findings. The most commonly selected image is the fish reflected in a vertical line, with placements shown in Figure 4b. This shows again how the horizontal and vertical dominate impressions of symmetry. Many students believe that objects that look horizontal or vertical always stay so under reflection. In fact over 90% of students selected one of the four "horizontal" images (first, second, fourth, fifth in Figure 4a). There is a strong 'gestalt' preserving the horizontal and vertical. Only 4% selected the correctly oriented fish (Figure 4b), and many fewer also positioned it adequately. The balance interpretation is evident in Figure 4c. The image selected is rotated by 180^{o} (upside down, turned fish), mostly positioned in the top left corner, 'balancing' the original in the right lower corner. Bell (1993) explained this as an association of "reflecting with various pairs of opposites such as forwards and backwards, towards and away, left and right, upwards and downwards" (p. 131).

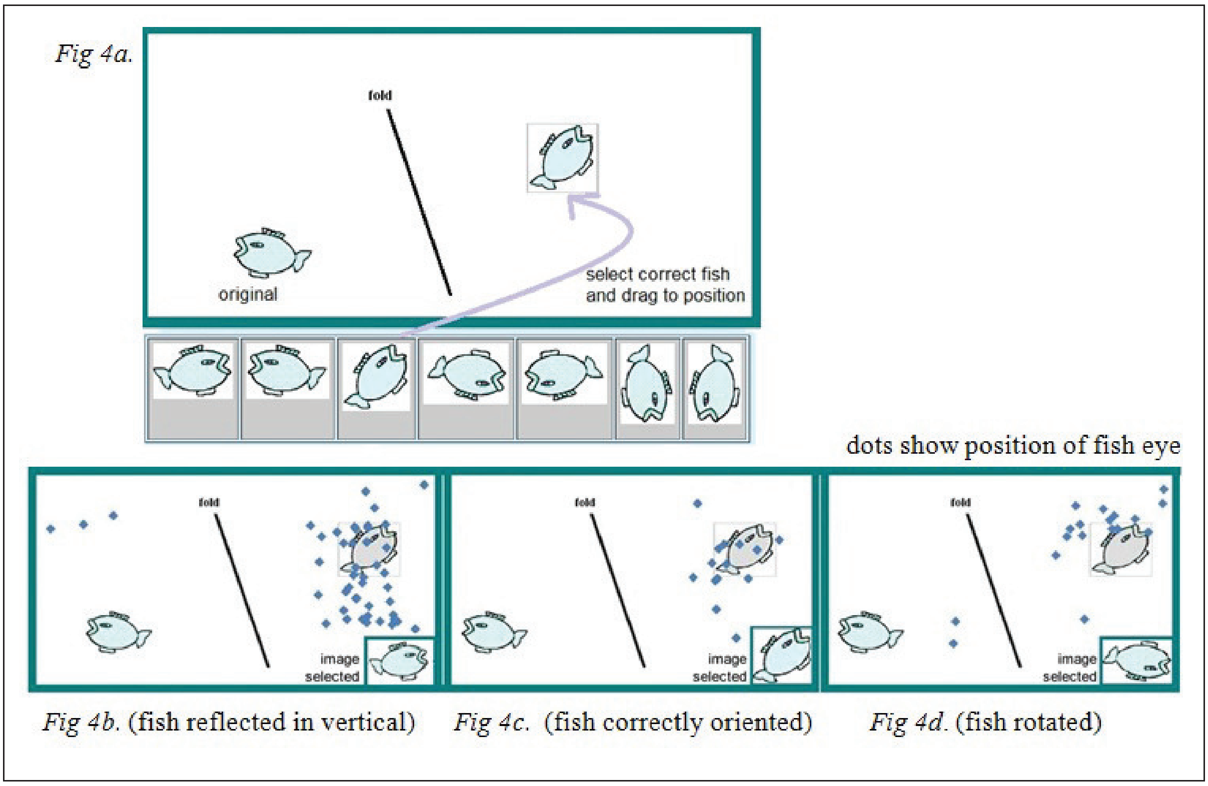

Figure 4. Reflect a complex shape: item and scatterplots for three oriented images.

From Reponses to Results: Evidence Identification and Accumulation

After students complete a test, the next process is evidence identification – what information can be drawn from the responses. For the Reflections test with its drag and drop format, the images selected and their exact placements can be automatically scored, giving correct/incorrect information, although setting realistic tolerances is tricky. Success or an error can be classified by automated checking in which nominated region an image is placed and how it is oriented. Crucially, for an understanding of students' thinking that can lead to better teaching, it is not just correct/incorrect that is important, but finding and then reporting the patterns in students' responses that reveal their thinking. This is a unique feature of the smart tests: the items are planned with this 'mapping of learning' in mind (Steinle & Stacey, 2006) and patterns in the responses are sought. The smart tests report in two ways, described below.

Developmental stages. There are many possibilities for reporting students' results to teachers. The approach selected for the smart tests is to describe learning in terms of stages along a developmental sequence, one sequence designed for each test, supplemented with additional information about common errors. Developmental stages highlight a small number of important underlying ideas, to address in teaching. This approach was discussed by Stacey, Price and Steinle (2012). The developmental stages for Reflections are shown in Table 1, with the percent of sample students at each stage in the final column. We often observe that the percentages in different classes vary markedly, so the averages are often not a good guide for the teaching of any one class. In this case, the first draft of the stages proposed that students would first be able to deal with simple shapes (e.g. blobs) with horizontal or vertical fold lines, then with oblique lines, and then master complex shapes. However, the stages below better fit the data. Stacey et al (2012) describe the data analysis processes. Stage descriptions need to be easy for teachers to understand and easily related to teaching actions since face-to-face professional development is not available on-line.

Table 1
Developmental stages for Reflections smart test (2013) with percent at each stage

Stage	Description		% at stage* (N = 504)
Stage 1	Students have a general but imprecise idea of reflection		35%
Stage 2	and	can accurately reflect simple and complex shapes in a horizontal or vertical line	34%
Stage 3	and	can reflect a simple shape (such as a circle) in any line, including oblique lines	9%
Stage 4	and	can reflect a complex shape in any line, including when the shape and the mirror are not visually aligned	1%

* 21% of students are below Stage 1

Identifying mistakes, missing knowledge and misconceptions. In addition to indicating a developmental stage, test responses often reveal why students do not perform at a higher level, and so smart tests also report students' systematic errors to teachers. Identifying systematic errors across multiple responses is interesting but can also be a tedious and complex task, so there are real advantages here in computer diagnosis. For example, students who reflect as if all fold lines are vertical or horizontal can be identified from their responses as in Figures 3a, 3b, 4b. Looking for patterns in responses across items requires significant evidence accumulation and so the smart tests need considerable complex programming operating in the background to make this possible. One of the many programming challenges for our work is to distinguish serious errors worth reporting from 'careless mistakes' caused only by inattention. As more data becomes available, we refine the criteria for deciding that an error is likely to be systematic and important, and which items contribute to its diagnosis. Teachers who know about likely misconceptions or other systematic errors can plan teaching to address or avoid them. If they have information about individual students in their own class, they can provide targeted remediation.

The systematic errors identified by the Reflections test relate to the horizontal-vertical dominance, the 'balance' positioning, and getting either the distance or orientation consistently wrong. As with all categorisations, somewhat arbitrary decisions need to be made about where to draw the line in reporting on such errors. For instance, the percentage of students reported as making 'balance' placements will vary widely depending on decisions about the placement tolerances allowed, the number of items over which consistency is sought, and unique characteristics of the items. For example, the percentage of balance placements for the item in Figure 3b was around 5% (504 students), but for another apparently similar ‘blob’ item it was around 30%. This is probably because item 3b was given in conjunction with another blob placement and those two blobs together stimulated a horizontal gestalt, which dominated the wrong answers, and consequently reduced the number of balance responses. Making decisions about what is useful to report needs extensive data, and a sense of what would be useful for teachers. Because it is formative, rather than summative assessment, it is better to over-report than under-report.

Implications for Teaching about Reflections

Taking an overview of this data provides important information for teaching. From their practical experiences at school in flipping and folding, it seems that many students have a general visual understanding of reflection (and line symmetry) but it is vague and they do not analyze the situations mathematically. Some do not appreciate that the folding process makes the distances from the fold line equal and why. Many only vaguely visualize folding when the fold line does not appear horizontal or vertical on the paper. In case it is thought that such fold lines do not occur in real life, consider those in Figure 5. Many students miss the flower’s oblique lines of symmetry, or exhibit the common misconception of assuming line symmetry when two halves of a shape are congruent.

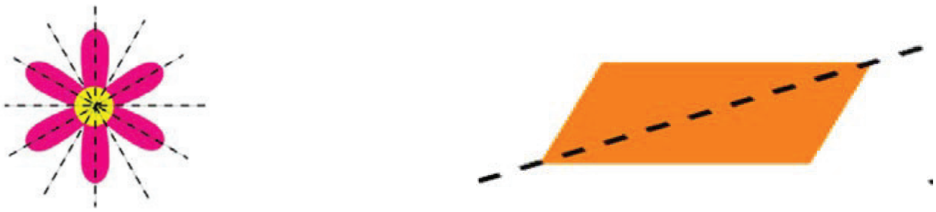

Figure 5. Lines of symmetry that are not horizontal or vertical are harder to identify.

The Australian Curriculum rightly identifies geometric transformations as a topic which lends itself to practical exploration, having strong links to the real world and with potential to enable students to be creative e.g. in making beautiful symmetric designs. It describes the learning experiences with verbs such as investigate, identify and create. What is less evident in the description is that symmetry is a real world phenomenon which has to be mathematized so that visual impressions form the foundation for abstraction and generalization with predictive power. Students must impose a precise mathematical lens on pictures and objects and actions. The missing activity verb from the Australian curriculum is measure: measuring lengths and measuring angles to discover and make precise the mathematical properties. There are subtleties in the move from the real world to the mathematized world. For example, manually flipping a concrete object (e.g. a cardboard shape) involves a physical rotation around a line through the third dimension. With concrete objects, reflection can only be achieved by rotating. Flipping is good for seeing the change in orientation under reflection, but it does not assist with the position of an image or any notion of a line of symmetry. On the other hand, from practical experiences measuring the effects of folding, students can begin to develop the concepts of object, image and line of symmetry and the basic mathematical rules. An important part of PCK is knowing how to make transitions from practical to mathematical experiences.

Further increasing teachers’ PCK. One function of smart tests is to demonstrate some of the dimensions of complexity which make mathematical tasks more challenging, and which need to be included in teaching: in this case, object complexity and the orientation of the fold line to the paper and of the object to the fold line. However, teaching advice needs to give information to teachers about other types of tasks for reflection, which get students to think in other ways: finding lines of symmetry in shapes such as in Figure 5, drawing the line of symmetry given the object and image, and combining transformations.

Conclusion

Smart tests provide a different type of formative assessment, aimed at activating research on s tudents' understanding for use in classrooms. Teachers can learn about individual students. Where results are puzzling, they can look through the individual students' responses. They can plan teaching directed more closely to the needs of the individual, group or the class. We also hope that the test items and the reporting of the results highlight for teachers demonstrate (part of) the range of tasks that needs to be encountered as students learn about a topic. As Agent Smith said in the film 'The Matrix': "Never send a human to do a machine's job." In formative assessment, we hope to be breaking new ground in what is best done by a teacher and what is best done by a machine.

Acknowledgements

I gratefully acknowledge the contributions of B. Price, E. Gvozdenko, V. Steinle and earlier H. Chick, initial funding from the Australian Research Council (LP0882176), the Victorian Department of Education and Early Childhood, the Catholic Education Office (Melbourne) and the teacher and students who contributed data.

References

Australian Curriculum and Reporting Authority (n.d.). *Australian Curriculum: Mathematics*. Retrieved 31 May 2013 from World Wide Web: http:// www.australiancurriculum.edu.au/Mathematics

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klausmann, U., Krauss, S., & Tsai, Y-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.

Bell, A. (1993). Some experiments in diagnostic teaching. *Educational Studies in Mathematics,* 24, 115-137.

Kűchemann, D. (1981). Cognitive Demand of Secondary School Items. *Educational Studies in Mathematics,* 12, 301 – 316.

Schultz, K. (1978) Variables influencing the difficulty of rigid transformation during the transition between the concrete and formal operations stages of cognitive development. In R. Lesh & D. Mieriewicz (Eds.), *Research Concerning the Development of Spatial and Geometrical Concepts*. Columbus, Ohio:ERIC

Stacey, K. & Steinle, V. (2006). A case of the inapplicability of the Rasch Model: Mapping conceptual learning. *Mathematics Education Research Journal*, 18(2), 77 – 92.

Stacey, K. Price, B., Steinle, V. (2012). Identifying stages in a learning hierarchy for use in formative assessment – the example of line graphs. I n J. Dindyal, Cheng L-P, Ng, S-F (Eds.) *Mathematics Education: Expanding Horizons* (Proceedings of the 35th Annual Conference of Mathematics Education Group of Australasia, pp. 393 - 400). Adelaide: MERGA.

Steinle, V. & Stacey, K. (2012) Teachers' Views of using an on-line, formative assessment system for Mathematics. *Pre-proceedings. 12th International Congress on Mathematical Education Topic Study Group 33*, 8– 15 July, (pp. 6721 - 6730). COEX, Seoul, Korea.

Steinle, V., Gvozdenko, E., Price, B., Stacey, K., & Pierce, R. (2009). Investigating Students' Numerical Misconceptions in Algebra. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides.* (Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia, pp. 491 – 498). Palmerston North, NZ: MERGA.

The Beginnings of MERGA

Preamble to the Annual Clements/Foyster Lecture

In the middle of 1976 John Foyster, who was then based at the Australian Council for Educational Research (ACER), came to see me at Monash University, where I was in charge of the Mathematics Education program. John talked about how the Australian Science Education Research Association (ASERA) had recently been established, with Professor Richard Tisher (then of Monash University) as the prime mover. John wondered whether the time was ripe for a similar national group interested in *mathematics* education research to be established, and asked whether he and I might take steps to establish such a group.

My immediate reaction was yes, we should do it. Then came the doubts and reservations. How would the Australian Association of Mathematics Teachers (AAMT) react to such an initiative? After all, AAMT already had a "Research Committee." In any case, would there be enough mathematics educators in Australia, interested in such a group to make it a viable proposition? Who would provide the funds likely to be needed for the establishment of such a group?

It was John's and my opinion that the AAMT Research Committee had not reached out to embrace most of the people lecturing in mathematics education in Australia at teachers colleges or in universities at that time. Intuitively, I thought Australia needed a group like the one John was proposing. My intuition told me that AAMT was not the organisation to move towards the establishment of such a group.

John assured me that he would put up any funds needed to get the group going (and, hopefully, any group that was established would be able to pay him back within a few years). Hence we decided to proceed with the idea of establishing the group and to strike while the iron was hot, so to speak, by conducting a national conference at Monash University in the middle of 1977. I came up with the name "Mathematics Education Research Group of Australia" which John liked because of the acronym MERGA, which suggested a "merging together." We sent out notices of our intention to form MERGA late in 1976. Neither of us knew many of the people who might be interested in joining such a group, so the notices were addressed to the "Mathematics Lecturers at …"

Soon after we had decided to go ahead, I heard of the existence of a group, based in New South Wales, called the Mathematics Education Lecturers' Association (MELA). John and I talked about whether MERGA and MELA might become one from the outset, but we decided that the aims of MELA seemed to be sufficiently different from those that we envisaged for MERGA, focused far more on research than lecturing, that we should proceed with the MERGA idea.

And so it came to be that in May 1977, the first of what was to become the annual conference of MERGA took place. About 100 people attended, with papers frenetically being read from 9 am to about 10 pm, for three days, in a Rotunda Theatre at Monash University. Professor Richard Tisher was present at the start of the Conference, and talked of his experiences in establishing ASERA. Frank Lester, of Indiana University, was among those present. In the event, two volumes of papers read at the Conference were produced (the first volume being available on the first day of the Conference, and the second several months later).

At a post-Conference meeting it was decided that, yes, MERGA should be formed, that the second meeting would be at Macquarie University in May 1978, and that an annual conferences should be held each year at a different academic institution. At that second conference it was decided by those present that MERGA should continue and a constitution and election of offices would be decided on at the third conference to be held at the then Brisbane College of Education. And so MERGA was born.

Ken Clements