

MATHSMAPS FOR DIAGNOSTIC ASSESSMENT WITH PRE-SERVICE TEACHERS: STORIES OF MATHEMATICAL KNOWLEDGE

Julie Ryan and Julian Williams

Manchester Metropolitan University and University of Manchester

**This is a copy of our paper accepted for publication in
Research in Mathematics Education, Volume 9 and offered
for discussion at the Nuffield Seminar Series on Mathematical Knowledge in
Teaching September 2007**

This paper reports an innovative assessment feedback tool – the mathsmap – and describes how two pre-service teachers made sense of this personalised diagnostic map to reflect on their own subject knowledge in mathematics. The mathsmap provided both a summative and a diagnostic profile of their attainment and errors across the mathematics curriculum required for teacher training. The use of the mathsmap to reflect learning on a personal level is seen to also provoke ‘accounts’ or ‘stories’ that might inform pedagogical content knowledge: in making their mathsmap comprehensible to themselves, the teachers needed to account for their own knowledge-troubles, that is, to narrate their metacognition.

INTRODUCTION

An important consideration in teacher education is the subject knowledge of trainees. The transformation of mathematics subject matter knowledge into pedagogical content knowledge is a significant step in teacher education. Subject matter knowledge is more than knowledge of facts or algorithms – it requires knowledge of both the substantive structure (facts and their organising principles) and the syntactic structure (legitimacy principles for the rules) of the subject domain (Goulding, Rowland & Barber, 2002; Rowland, Martyn, Barber & Heal, 2001).

The required minimum level of school mathematics achievement for entry to primary teacher education and non-mathematics specialist courses in England is typically GCSE grade C. However, a GCSE attainment level does not provide fine detail about subject matter knowledge. The *Teacher Education Mathematics Test (TEMT)* (ACER, 2004) developed in Australia and used with pre-service teachers in England was developed to provide summative and diagnostic assessments of individual attainment across the mathematics curriculum including both substantive and syntactic understanding (Ryan & McCrae, 2005; Ryan & McCrae, 2006).

The diagnostic component of the test reported the errors made by the pre-service teachers in order to inform either personal development or collective treatment within a cohort. Teacher errors deserve attention not least to avoid transfer to children in schools. Moreover we suggest that errors provide positive opportunities for pre-service teachers to examine the basis of their own understandings and promote a pedagogical strategy for their own classrooms. If they are to learn to treat their learners’ errors with respect and engagement, then they must come to so value and engage with their own.

The personalised map of test response indicates the secure and non-secure curriculum areas of individual pre-service teachers: it indicates ‘gaps’ in knowledge or faulty conceptions in terms of expected outcome given the pre-service teacher’s summative attainment level.

TEST DEVELOPMENT

The *Teacher Education Mathematics Test (TEMT)* (ACER, 2004) was developed by first constructing a ‘teacher curriculum’ based on Australian and UK curriculum documents with the level of attainment targeted at Australia’s school level 5/6 which is the equivalent of England’s GCSE grade C.

A bank of 105 multiple-choice pen and paper items was then written to the curriculum and trialled. Calculators were not to be used as written computation was considered to be fundamental subject knowledge – other non-number items were written to be independent of computational skills. Three tests of 45 items were constructed with 15 common (link) items distributed within the first half of the test and in the same location. The tests were timed for a 45-minute testing period. The six curriculum strands covered were: Number (16 items in each test), Measurement (8), Space and Shape (8), Chance and Data (6), Algebra (5), and Reasoning and Proof (2). Marks were not deducted for incorrect responses. (For further details see Ryan and McCrae, 2005).

The *TEMT* items were also written with diagnostic coding for most distracters (three or four per item). A range of mathematics education research on children’s and teachers’ knowledge and errors informed the writing of the *TEMT* items and choice of distracters. It was also seen to be important to provide adult contexts for test items and to take advantage of the presumed higher reading ability of adult students.

Trainees across three different degree courses (total $N = 426$) took a *TEMT* in the first few weeks of the first year of their primary teacher education degree at a university in Australia in 2004. The three test versions were trialled.¹

A Rasch analysis (Bond & Fox, 2001; Rasch, 1980) was undertaken using Quest software (Adams & Khoo, 1996).² Quest provides classical statistics as well as item estimates (item difficulty estimates with the mean difficulty set at zero), case estimates (student ability) and fit statistics. Each test item is *scaled* in terms of its difficulty (usually from -5 to 5 logits) and each trainee is located on the same scale in terms of their ability as measured by the test. The data were found to be compatible with the Rasch model. The three test forms were found to be equivalent. Test reliability and goodness of fit were strong and are reported in detail in Ryan and McCrae (2005).

A second cohort of pre-service teachers in England (total $N = 87$) also took a *TEMT* assessment (the same test form) in the second year of their training in 2005. Their patterns of response were very similar to the Australian sample.³ These pre-service teachers included primary trainees, non-mathematics specialist secondary trainees and a small group of mathematics secondary trainees; participation was on a voluntary basis with the promise of personalised diagnostic feedback from the test to assist their subject knowledge development.

The 87 trainees in England were then given an individual map of their responses as diagnostic feedback. A questionnaire gathered information on what sense they made of their map and two pre-

service teachers from this cohort were interviewed further to see what sense they made of this feedback and how they intended to address their indicated mathematical needs.

PERSONALISED DIAGNOSTIC MAP

Quest software produces a *kidmap* (here called a *mathsmap*) that is an output for each individual highlighting their correct *and* incorrect response patterns. The map summarises an individual's performance according to the Rasch model's expectations.

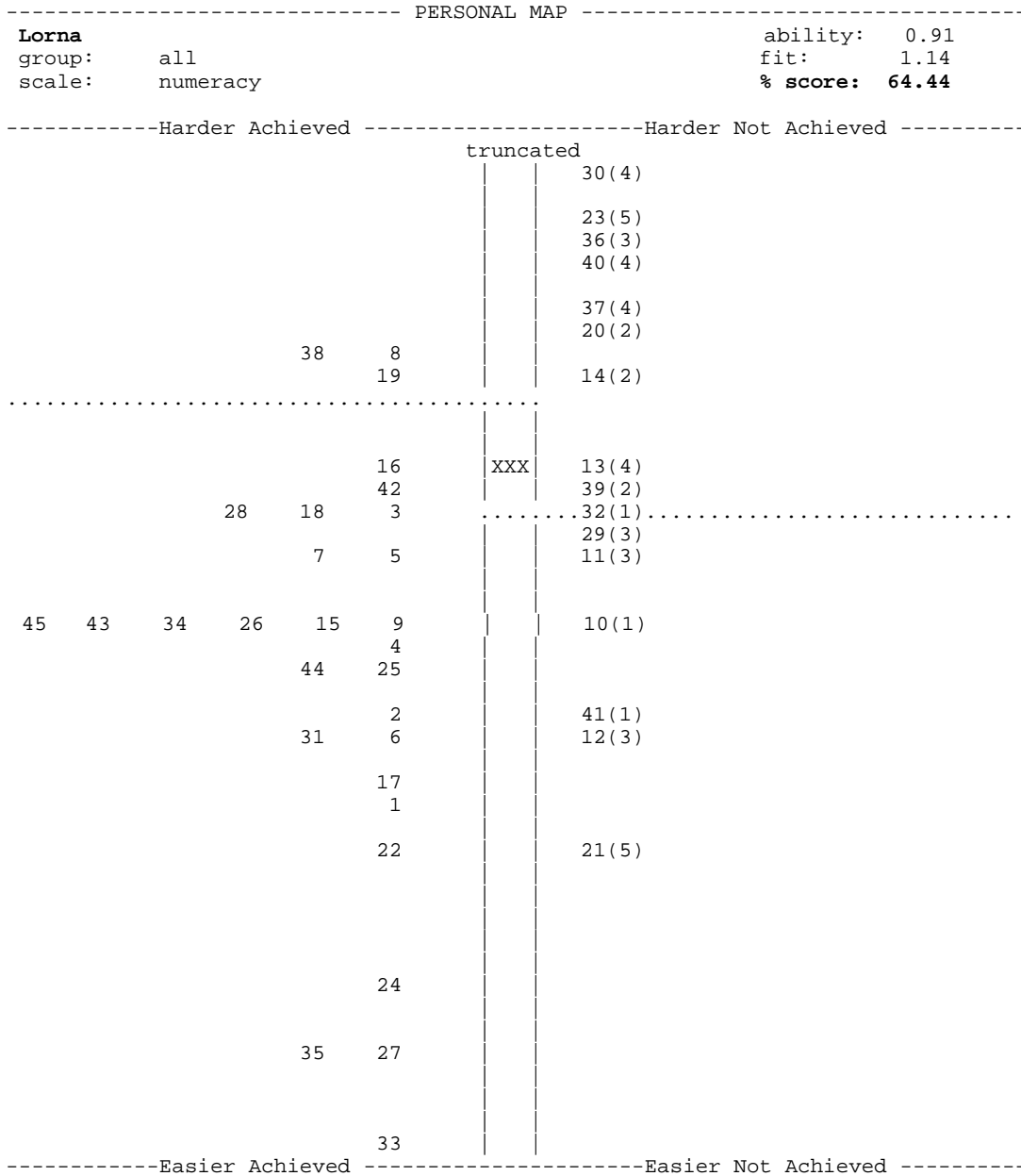


Figure 1. Mathsmap for Lorna

The map locates each test item on a vertical scale according to its difficulty (easiest up to hardest) and then separates items horizontally (left or right), according to whether the student answered them correctly or not. The map also locates the individual according to ability on the same vertical

scale (centrally marked with 3Xs). Under the model an individual has an increasing probability of achieving items below their ability estimate and a diminishing probability for achieving items above their estimate.

The items achieved by the individual are plotted on the left-hand side and the items *not* achieved are plotted on the right-hand side of the map. Additionally the actual option choices made for each incorrect item on the right-hand side are indicated in parentheses.

The student ability estimate is located on the vertical scale and the student's 'fit' to the model (infit mean square value) is reported in the print-out.⁴ An example is shown in Figure 1 where Lorna has an ability estimate of 0.91, a mean square infit statistic of 1.14 and a total score of 64% (29 of 45 items correct).⁵ The row of Xs (centre of the map) locates her ability estimate (0.91 in this case) and the dotted lines represent ± 1 standard error for the estimate.

The individual would be expected to achieve all the items at and below their ability estimate with an increasing probability for those further below. Lorna has a 50 percent probability of answering items *at* her ability estimate (note that item 16 is correct and item 13 is incorrect). She would have been expected to have achieved items 39, 32, 29 and so on, but she answered incorrectly with options 2, 1 and 3 respectively (shown in parentheses on the right). Lorna would not have been expected to correctly answer items 38, 8 and 19 which are located *above* her estimate (on the left) but she did respond correctly. In a perfect 'goodness of fit' to the Rasch model, the top left and bottom right quadrants would be empty so items in these quadrants are particularly compelling for discussion in the first instance.

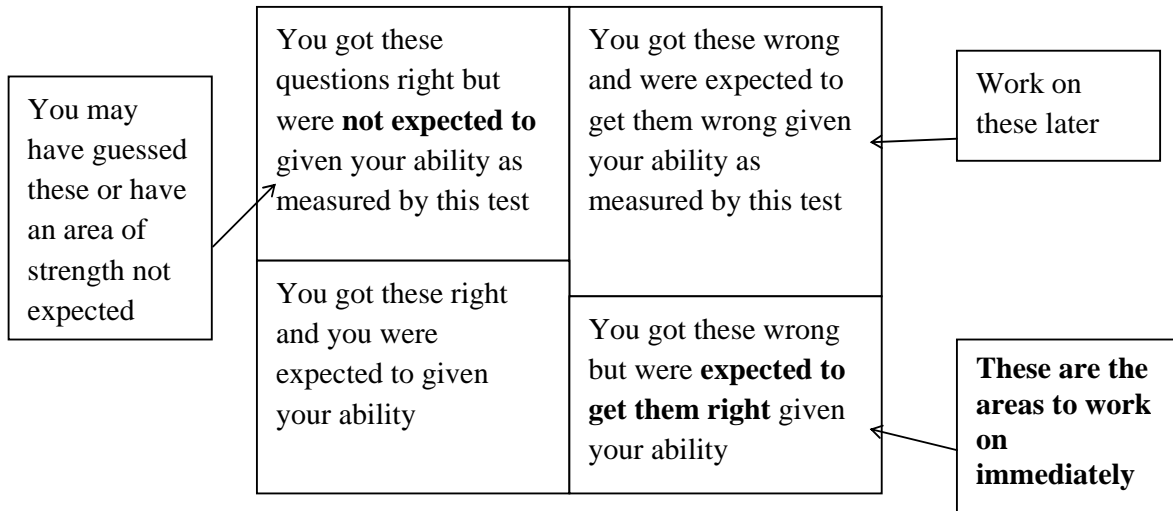


Figure 2. Summary of 'How to read your mathsmat'

The trainees were given their individual map, an instruction sheet on how to read it (see Figure 2 for a summary diagram from that information) and a list of the test item descriptors only. The test items were withheld so that the curriculum area indicated by the descriptor was targeted for study by the trainee in a broad sense rather than in terms of item-specificity. See Table 1 for Lorna's 'easier not achieved' item descriptors.

Item	Curriculum Description of Item
29	Algebra: multiplying simple algebraic expressions by a number
11	Chance: likelihood/probability of everyday events (numerical)
10	Shape and Space: identifying Cartesian co-ordinates
41	Algebra: from tables of values to algebraic rule
12	Chance: recognising dependent events (reduced sample)
21	Measures: finding perimeter of a rectangle – words

Table 1. Descriptors for Lorna’s unexpected *incorrect* items

TRAINEE USE OF THE MATHSMAP

Lorna (ability estimate 0.91) and Charlene (ability estimate 2.00) agreed to be interviewed on how they used their mathsmap and interpreted the accompanying explanatory documents.⁶ They sat the same form of the *TEMT* test but had different profiles in terms of mathematical confidence, experience and teaching practice. We outline here their reactions to their maps, accounts of their errors, study techniques and their responses to particular test items.

Lorna

Lorna was a mature aged trainee studying on a 4-year BA Primary (Hons) with QTS course (enabling her to teach in primary schools). She was not confident about her mathematics ability and said that she had achieved a C grade in mathematics in O-levels some 20 years ago. However she had answered 64 percent of the items correctly and was interested in targeting areas of weakness in her subject knowledge (see Figure 1 for her mathsmap).

Lorna: [The map] was a little confusing at first, but I soon got the hang of how to read it with the help you sent. It identified areas I thought I was weak in and some I didn’t.

The items in the top left quadrant of a mathsmap are the ‘harder achieved’ items. Unexpectedly correct responses may diagnose guessing in any multiple-choice test format.⁷ However, Lorna reported *recent targeting* of the curriculum area indicated in the top left quadrant of her mathsmap while on teaching practice because she already knew this was an area of weakness – she had not guessed here. Her items in this quadrant were all Shape and Space (see Table 2 for descriptors of items 38, 8 and 19).

Lorna: Well that’s interesting, that! Because on my teaching practice last year with year 6, I did a unit of work in term 1 for Shape and Space and it was all about quadrilaterals and rotating shapes and the size of angles (and) symmetry. So maybe that is where that has come from, that not only I have taught them but I have learnt as well ... So I have ... as well as teaching children I have learned myself, so I know I have learnt more from what I have taught, as well as teaching at the same

time ... (Excited) so that tells me that maybe with time and practice that this area here [bottom right quadrant] will come, up ... over.

Lorna seemed pleased that her mathsmap suggested that she had already successfully targeted a shaky area of curriculum and she commented on her subject knowledge improving as she taught. After an unsuccessful lesson, her school mentor had given her time to study and prepare the lessons for this area again, so Lorna had collected textbooks and had used the internet to study Shape and Space extensively on her own in order to feel more confident.

Lorna: (I used the book on) subject knowledge, it's the one we have here in the library. And I went out and bought it and I just sat and read and read and read on Shape and Space. ... I think it's by Suggate. ... It was in the directed reading notes we were given to do every week. I went to that one because I'd done the (chapter) on algebra, because I was rusty on algebra. So I read up on algebra and found it really useful. It worked for me. The vocabulary was good for me. So I thought, right, I'll go for it and use it for Shape and Space. And obviously it did, it helped, it worked. I thought, now I know what to do and I went out and bought it.

Item	Curriculum Description of Item
38	Shape & Space: rotation of a shape about an internal point
8	Shape & Space: interpreting drawings on a grid
19	Shape & Space: finding one missing length for similar shapes

Table 2. Descriptors for Lorna's unexpected *correct* items

She then referred to the items in the bottom right quadrant – 'easier not achieved' which she now felt she could be successful with using the same study strategy.

Lorna: It shows me that there are a lot of concepts there that are quite rusty because I am 39 – (that's) 20 years after [my own schooling] ... so that tells me that maybe through teaching that I, (with) just a little bit of homework and practice, that I could move those quite easily up ... over, to there [left]. ... Because I do *fear* maths, I see maths as a bully. It is *my* bully. And this has shown me that I can overcome this, and become an effective maths teacher.

Lorna also identified Algebra as one of her "rusty" areas and was becoming confident that she could move it 'over the line'. She asked to discuss an actual test item. Her discussion of item 41 (see Table 1) showed that she could now talk her way through the item on matching a table to an algebraic rule (see Figure 3) after having done some personal study on Algebra.

Which of the following tables represent the function $y = x^2 + 3$?

Table 1

x	0	1	5	10
y	3	4	8	13

Table 2

x	0	1	2	5
y	3	4	7	28

Table 3

x	0	1	2	3
y	3	5	7	9

A. Table 1 only B. Table 2 only C. Table 3 only D. Tables 1 and 3 E. Tables 2 and 3

Figure 3. Item 41: 'Algebra: from tables of values to algebraic rule'

Lorna: Question 41. (Looking at her test script) I wrote at the side 'guessed, no idea!'

Interviewer: Do you want to talk through now what you are thinking perhaps?

Lorna: First thought, ooh, algebra! Right! So, you've got to work out – I can graph this scale, if x is squared plus 3, you are going to have a plus – you're going to have it going plus 3 every time *but* it's got to be squared as well. So you're going to have to take 3 off, and then you've got to have a number that you can get a square root from. This is *after* now reading about algebra. *Before* I would have just thought, oh, well it must start with a 3. And then I've thought, no, hang on, how am I going to do this? I just didn't know. And then I thought, oh x , in the top row in table 1, you've got 1, then I felt, well 'x squared', 1 times 1 is 1 plus 3 is 4 (pointing to it)... And then the next number along in table 1, x . I've thought if x is 5, I've not squared it, I've just added 3. And the next one along in table 1 is x is 10, and then the answer below is 13. I've just added 3, I've just guessed, *panicked* and just gone for number 1 [option A] which was table 1.

Here Lorna constructs an account of her mistake of 'adding 3', which she had originally thought was because she "guessed": she now 'after reading about algebra', can see " x is squared plus 3 ... you're going to have it going plus 3 every time *but* it's got to be squared as well". She reinforces this formulation of the function by inverting it and emphasising the need for a square root.

We note that in talking about her own thinking 'before', she switches tenses as in "I would just have thought" and "I just didn't know". Here she constructs her old thinking to include a squaring of the x , re-working the first x -value in Table 1, getting the right value of 4; but then "I've thought, if x is 5, I've not squared it, I've just added 3 ... I've just added 3, I've just guessed".

What began with a "guessed, no idea" becomes, by the end of her story, a new guess, "I've just added 3" which we pedagogues would conceptualise as a self-diagnosis. This is an important storying of her self 'before' and 'after' her learning about algebra, and we think offers insight into her potential metacognitive learning about her own learning.

Charlene

Charlene was a science specialist trainee on a 3-year BSc (Hons) in Primary and Secondary Education with QTS (enabling her to teach as a generalist in KS2 or as a science specialist in KS3 and perhaps KS4). She was confident in her mathematics ability – she had answered 80 percent of the items correctly and was interested in seeing where she had made mistakes. She had achieved a B grade on her AS- level mathematics two years previously. She reported that her mathsmat (see Figure 4) was initially a puzzle but once she had read the detailed instructions it made sense.

Charlene: When I first looked at it, I was like 'what is this!' I was looking at it thinking 'how do you read that?' But then, once I'd ... actually looked at it properly, and then

read a few of the instructions, I was like ‘that’s easy!’, it made sense, and it seemed the best way, probably, to present the information.

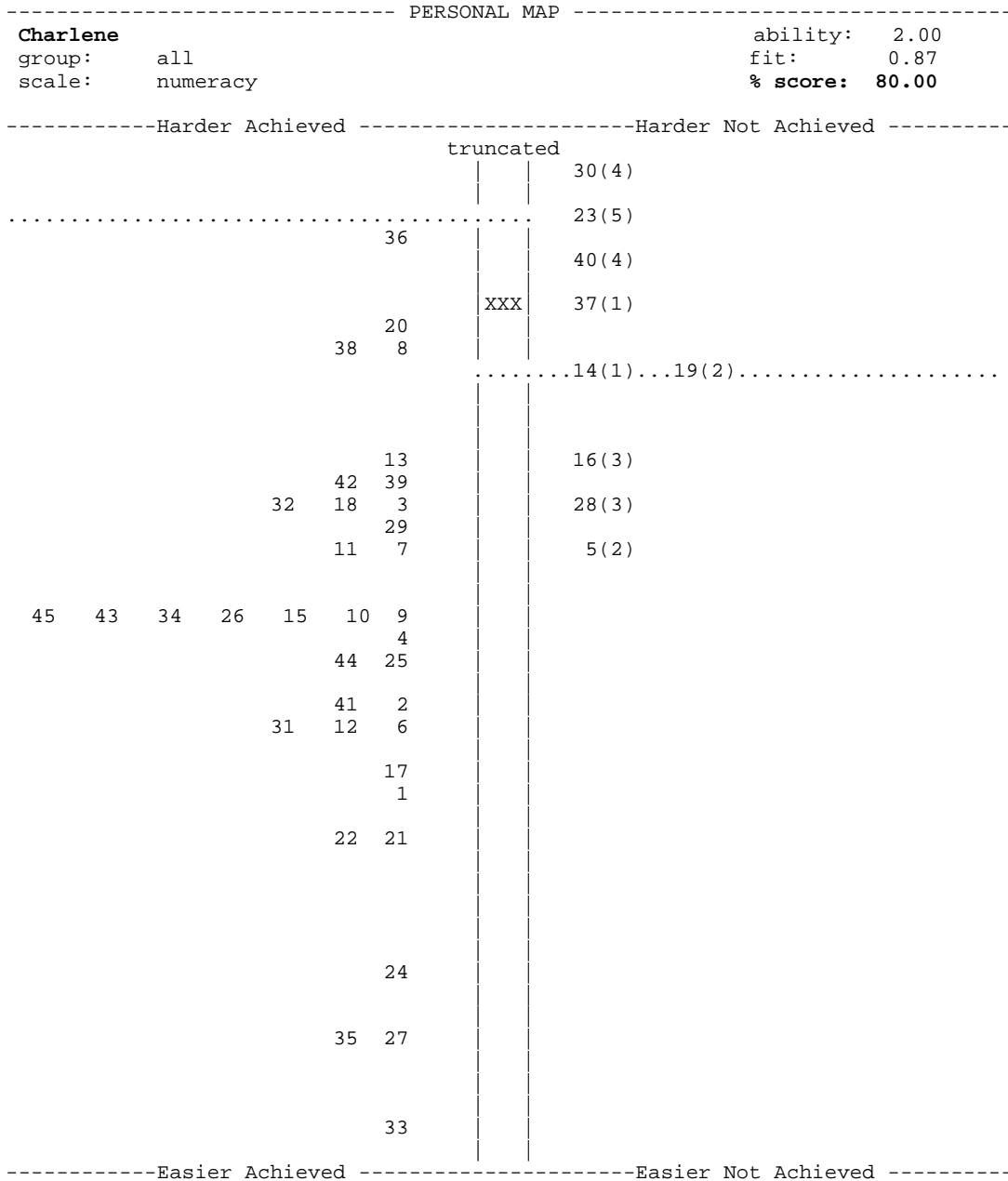


Figure 4. Mathsmap for Charlene

Charlene had systematically matched the questions by colour-coding the curriculum descriptors and the items on her mathsmap right-hand quadrants (see Figure 4 for her mathsmap). She confirmed that the items in the bottom right quadrant made sense as items she should have answered correctly and seemed to have an understanding of the type of errors she would have made.

Charlene: I mean, they looked like the sort of things that I ... probably would have had problems with or made a silly mistake on, like the decimal point (question 16)... and also probably (question) 5 because it's 'measuring, in lengths, mm, cm and metres' so that will be converting, which is easy for me to make a mistake in. ... I just, I don't know, I just get carried away. I jump one step ahead, and it all goes

pear-shaped... ‘Cos sometimes I try and think too advanced for the questions, ‘cos I did AS [A- level year 1], not very well, maths, but I do sometimes think there’s more to it than what’s there.

Charlene suggests that she “get(s) carried away”, or thinks in a “too advanced” manner rather than having missing knowledge, that may explain her errors.

Item	Curriculum Description of Item
16	Number: Decimal to fraction conversion
28	Data: graphs – generating rules of the form $y = mx+c$ from graph points
5	Measures: ordering metric lengths stated in mm, cm, m

Table 3. Descriptors for Charlene’s unexpected *incorrect* items

Charlene reported that in converting 0.125 to a fraction on the test (item 16, see table 3) she probably ‘misread’ one of the answer options (C: 125/100) which she had selected thinking it was 125/1000 (not an option). But she also said that her mental mathematics skills needed improvement and her processing on this item showed that she was using repeated addition to find how many 125s in 1000.

Charlene: (Reading the question) “0.125 is the same as” (Pause) It’s, not sure how to do, it’s 1, 2, 5 over a thousand. I think I probably went for C originally. (Checks) Yes... Because I just must have missed out, misread one of the noughts, seeing there was an extra nought on it, because that was an automatic ...

Interviewer: What would you go for now?

Charlene: (Long pause) I need to improve my mental maths. I can’t. (Pause) I’ll have to do it the long way...

Interviewer: What’s the long way?

Charlene: (Laughs) I’m doing, how many, I’m working out the multiples of 125, to work out whereabouts (writing) a thousand ...

Interviewer: You’ve got 125, 250.

Charlene: 375, 500. OK, so 4 is 500, so, 8 would be a thousand. So it’s ‘1 over 8’, which is B.

Interviewer: You’ve gone for B. So why do you think you went for C originally, again, can you express that?

Charlene: Because I misread the 100 as 1000, so I just assumed it was 125 over 1000 when it was 125 over 100. And I think even when we came out, somebody mentioned that, and I thought, oops, maybe I did pick the wrong one then.

This account matches Charlene’s first explanation for her ‘mistakes’ as getting “carried away” or “jumping ahead” so that things go “pear-shaped”: she said she “misread” and ‘saw’ an extra nought in the denominator of the option C fraction and processed quickly here as a one-step item. Here for

item 16, her thinking does not appear to be “too advanced” or anticipate a two-step item, but rather suggests a seldom-used mental fact which took her a little time to re-construct.

One of Charlene’s items located *at* her ability level (see Figure 4, item 37) was answered incorrectly. The curriculum description was ‘dissection and tessellation: understanding Pythagoras’ theorem’ and involved interpreting a classic proof by area dissection (see Figure 5). It was the fifth hardest item on the test but discriminated well at the top end of the ability range. Charlene said it was an unusual question because it was asking for a proof.

An internet animation demonstrates the theorem of Pythagoras by dissection and *drag-and-drop* transformations of the shapes shown on the diagram.

What will the transformations show to demonstrate the theorem?

- A. That D and C will fit exactly into E
- B. That A, B, C, D and E will fit exactly into F
- C. That A, B, C, D and G will fit exactly into F
- D. That A, B and C will fit exactly into G

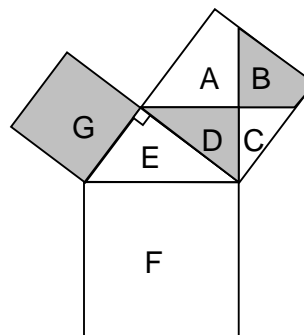


Figure 5. Pythagoras’ Theorem

Charlene: (Laughs while reading the question) No, it’s just, yes, what’s this on about? I think it could just be the question itself as well, (if) you’ve not really experienced that sort of thing... It’s something that’s got to prove Pythagoras’ theorem and that ... Is that ‘ a squared plus b squared equals c squared’? Is that Pythagoras?

Interviewer: Is it?

Charlene: (Pause) I don’t..., or is it *sohcah*...No, *sohcahtoa* is different. It is ‘ a squared plus b squared equals c squared’. (Pauses)

Interviewer: What would that mean in relation to this picture?

Charlene: (Pauses and laughs) I haven’t got a clue! (Pauses) I don’t know what it means, the diagram... ‘ a squared plus b squared equals ...’

Interviewer: What does that mean?

Charlene: It means the length of the two short sides, both squared, and added together, is the same as the length of the longest side, the hypotenuse, squared ... (pauses)

Charlene juggled good-naturedly with the item here saying “what’s this on about?” and recognised that ‘previous experience’ of something like this would help – it was an unexpected type of test question. She ‘knew’ the Pythagorean theorem but appeared not to have a geometrical image of it and did not make any connection with ‘square’ shapes in this or further discussion – this is not surprising of course if the theorem is simply represented as a numerical/algebraic formula without visualisation. But the point is that she does not consider this as an instance of a missing conception of ‘square’.

COMPARISON AND CONTRAST

Lorna and Charlene had very different mathematical backgrounds, levels of confidence and motivation to improve their subject knowledge. As a mature aged student, Lorna was highly motivated and aware of her “rusty” knowledge and particular areas of weakness. She had in fact underestimated her mathematical ability as measured on this test and was actually above average for her cohort – she had thought she had “mountains to climb”. As a result of uncomfortable exposure of poor subject knowledge on her own school teaching practice, she had already targeted Shape and Space for study and was very pleased that her mathsmap indicated that she had achieved beyond her current expected ability level. It appeared she was also very motivated by her school mentor who had given her the opportunity to “start again”. She was very independent and willing to put in a lot of extra time – she commented that the younger students wanted it all done for them. Lorna had targeted algebra from her mathsmap for personal study already and demonstrated in discussion that her confidence in articulating algebraic structure was growing. She seemed to be very positive about the sort of feedback the mathsmap gave her and considered her subject knowledge as a ‘work in progress’.

Charlene had recently completed AS-level and was a high achieving science student. She was very confident about her mathematics ability and had quickly made sense of her mathsmap. She did not identify any areas of subject knowledge weakness and generally explained most of her errors as simple processing errors due to her tendency to rush or to anticipate questions as more complex than they were. This seemed to be generally the case from discussion of her errors though she exhibited some fundamental scale misconceptions related to linear graphing, for example, in item 28 with prototypical misreading of the scale. She did not appear to be alert to multi-step questions though she could identify them in discussion afterwards. Charlene did note that her mental mathematics skills needed further work, but predominantly diagnosed her errors as ‘slips’, and her narrative leaves little space for knowledge gaps or misconceptions. Indeed she said she would prefer to have the actual test questions back to review to see whether she had just made a silly mistake or whether she did not actually understand something.

In both cases, the limitations of the mathsmap as a tool become apparent. Firstly, it was fortunate that Lorna was able to identify Shape and Space as a topic strength but it is not particularly well-designed to profile topic strength being a short, item-focussed diagnostic tool. Secondly, Charlene being a high-scorer receives less diagnostic feedback than Lorna. The mathsmap, if it were computer adaptive, would avoid this, as items would be targeted at her ability. Finally, for the same reason, we might expect a particularly weak student to get less value out the mapping tool as currently designed.

CONCLUDING DISCUSSION

In previous work we and others have shown how teacher errors can provide opportunities for pre-service teachers to examine the basis for their own understandings, as well as identifying areas for attention by teacher educators (for example, Rowland et al, 2001; Ryan and Williams, 2007). We here offered one method for encouraging teacher reflection by having pre-service teachers personally confront their responses, errors and misconceptions with a mathsmap. However, the mathsmap is different from other feedback devices in drawing attention to non-normative responses

of the two kinds. Being told that responses are not ‘expected’ causes dissonance, or ‘trouble’ to be explained, such troubles generate ‘accounts’ or stories to be narrated to account for them (Bruner, 1996). These accounts – it seems to us – provide the researcher or teacher educator with some insight into the students’ self-knowledge, indeed their metacognitive knowledge, and even the students’ sense of self-efficacy or agency in their own learning.

Thus, Lorna narrates her unexpectedly correct responses with a story of her growth in competence and confidence in her capacity to learn. It is difficult not to interpret this as a very positive indicator. On the other hand, although (or just possibly because) Charlene was a higher scorer, her accounts for her unexpected errors tell a story of ‘slips’, and tend to marginalise explanations that might invoke her need to learn or fill knowledge gaps. We do not want to over-interpret these two limited cases, but rather point to the way ‘accounting for the unexpected’ in both cases impels a story of themselves as learners or mathematicians. The resources they use – for example, whether they invoke ‘misconceptions’ or not – reflect their metacognitive knowledge of learning and hence taps their pedagogical content knowledge. Interestingly, recent work asking Primary teachers to account for the unexpected errors of their children (as produced on the children’s mathsmaps) have similarly provoked accounts from their teachers, which draw on explanations such as ‘slips’ or ‘we’ve done a lot of that recently’ (Petridou & Williams, 2007). This leads us to propose that the mathsmap is a tool for provoking students to ‘story’ their own learning and knowledge, and hence becomes a diagnostic of their cultural models for narrating stories of ‘learning’ in general.

REFERENCES

ACER (2004) *Teacher Education Mathematics Test (TEMT)*. Melbourne: Australian Council for Educational Research.

Adams, R. J. & Khoo, S-T. (1996) *ACER Quest: The Interactive Test Analysis System*. Melbourne: Australian Council for Educational Research.

Bond, T. G. & Fox, C. M. (2001) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum.

Bruner, J. S. (1996) *The Culture of Education*. Harvard University Press.

Goulding, M., Rowland, T. & Barber, P. (2002) Does it matter? Primary teacher trainees’ subject knowledge in mathematics. *British Educational Research Journal*, 28 (5), 689-704.

Petridou, A. & Williams, J. (2007) Accounting for aberrant test response patterns using multilevel models. *Journal of Educational Measurement*, 44(3), 227-247.

Rasch, G. (1980) *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.

Rowland, T., Martyn, S., Barber, P. & Heal, C. (2001) Investigating the mathematics subject matter knowledge of pre-service elementary school teachers. In M. van den Heuvel-Panhuizen (Ed.) *Proceedings of 25th Annual Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 4, pp. 121-128). University of Utrecht: The Netherlands.

Ryan, J. & McCrae, B. (2005) Subject matter knowledge: errors and misconceptions of beginning pre-service teachers. In P. Clarkson, A. Downtown, D. Gronn, M. Horne, A. McDonough, R.

Pierce, & A. Roche (Eds), *Building Connections: Research, Theory and Practice: Proceedings of the 28th Annual Conference of the Mathematics Education Research Group of Australasia*, (Vol. 2. pp. 641-648). Melbourne: Deakin University.

Ryan, J. & McCrae, B. (2006) Assessing pre-service teachers' mathematics subject knowledge. *Mathematics Teacher Education and Development*, 7, 72-89.

Ryan, J. & Williams, J. (2000) National testing and the improvement of classroom teaching: can they co-exist? *British Educational Research Journal*, 26(1), 49-73.

Ryan, J. & Williams, J. (2005). Rasch modelling in test development, evaluation, equating and item banking. In T. Palm (Ed), *Proceedings of the Third International SweMaS Conference Umeå, October 4-15, 2003*, (Em. No. 48, pp. 11-31). Report Series of the Department of Educational Measurement: Umeå University, Sweden.

Ryan, J. & Williams, J. (2007) *Children's Mathematics 4-15: Learning from Errors and Misconceptions*. Open University Press.

Williams, J., Wo, L. & Lewis, S. (2007, this volume) Mathematics progression 5-14: Curriculum/age and test year effects. *Research in Mathematics Education*, 9.

¹ The 15 link items were taken by all 426 students and the other 90 items in the bank were distributed across the three versions and taken by 140, 142 and 144 students respectively. For further detail see Ryan & McCrae (2005).

² For further discussion of Rasch analysis see Ryan & Williams (2005), Williams & Ryan (2000), and Williams, Wo & Lewis (2007, this volume).

³ Australian sample (N=426) case estimate statistics were: mean of 0.64, standard deviation of 1.14 and reliability was 0.88; English sample (N=87) case estimate statistics were: mean of 0.85, standard deviation of 1.15 and reliability was 0.87. The 87 students in the English sample took the same version of the test. The pattern of item difficulties was similar to the pattern for the 426 Australian students, that is, the items were distributed similarly along the scale.

⁴ The infit mean square is one index of the 'fit' of the person's responses to the Rasch model's expectation. The expected value is 1. A reasonable infit here is between 0.7 and 1.3 (Bond & Fox, 2001, p.179).

⁵ Lorna was in the 56th percentile for the students taking the test.

⁶ Lorna and Charlene are pseudonyms. Charlene answered 36 of the 45 items correctly. She was in the 86th percentile for the students taking the test.

⁷ In this test construction, distracters were chosen primarily for pedagogical reasons (known or suspected errors) and a sufficient *number* of them (usually 3 to 4) were used to mitigate against guessing. This does not of course eliminate guessing as a phenomenon.