Assessing Pre-Service Teachers' Mathematics Subject Knowledge

Julie Ryan
Manchester Metropolitan University

Barry McCrae
Australian Council for Educational Research and University of Melbourne

We report the development of an assessment instrument that provides a profile of the attainment and the errors of pre-service primary teachers across the mathematics curriculum. We describe test development, analyses and test validation involving a sample of 426 pre-service teachers in the first year of their training in primary education courses in Australia. We discuss a range of errors, strategies and misconceptions made across different strands of the mathematics curriculum and show that pre-service teachers are making the same errors as children. A second sample of 86 pre-service teachers in England was used to validate the test. We also describe how these pre-service teachers in the second year of their program made sense of their personalised diagnostic profile from the test in order to develop their mathematics subject knowledge.

In Australia and England, students seeking admission to primary teacher education courses (both undergraduate and post-graduate) come with a variety of mathematical backgrounds. Since primary teachers are required to teach mathematics, the mathematical attainment level of these students is of interest in university admission decisions. However, the range of students' mathematical credentials prior to admission to teacher education courses makes informed selection difficult.

Many factors need to be considered in selection: Which mathematics subject did the student study in school, and to what level and how long is it since they studied that mathematics? Are there potentially strong teachers who may not have taken traditional routes in the school curriculum? Additionally, a single achievement grade provides no detail of student areas of strength or weakness. Evidence of mathematical attainment, thus, is currently often weak. We sought to strengthen the availability of evidence by developing an entrance test that can be used to *profile* the mathematical subject knowledge of beginning pre-service teachers. The test can also be used during a course to profile the development needs of particular pre-service teachers.

The Teacher Education Mathematics Test (TEMT) (Australian Council for Educational Research (ACER), 2004) is designed to test not only the mathematical attainment of pre-service primary teachers but also to identify their errors, misconceptions and strategies; that is, it has both summative and formative capability. The test development aim was two-fold: (i) to create a bank of valid and scaled multiple-choice items so that custom-made tests could be constructed of various difficulties and foci; and (ii) to create and validate test items that provide formative feedback, through the reporting of errors with diagnostic value.

We regard the errors and misconceptions of pre-service teachers and children alike as positive indicators of learning opportunities, and believe that pre-service teachers, if they are to learn to treat their learners' errors with respect and engagement, must come to value and engage with their own errors and misconceptions. Accordingly, we conclude our report with a discussion of how a group of pre-service teachers made sense of personalised diagnostic feedback in order to strengthen their mathematics subject knowledge during their teacher education course.

Development of the Test

Content: A Curriculum for Teacher Subject Matter Knowledge

A 'primary teacher curriculum' was first constructed from a consideration of the Victorian (State) *Curriculum and Standards Framework (CSF)* (Board of Studies, 1995; 2000), *Mathematics – a Curriculum Profile for Australian Schools* (*Curriculum Corporation, 1994),* and England's *Initial Teacher Training National Curriculum* (Department for Education and Employment, 1998; Teacher Training Agency, 2003). The test assumes level 5/6 attainment on the CSF in Victoria, which is the equivalent of grade C at GCSE in England. According to the CSF (p. 7), "It is expected that the majority of students will be demonstrating achievement at level 6 by the end of Year 10 – a realistic requirement for functional numeracy." The constructed curriculum covered the following six strands: Number, Measurement, Space & Shape, Chance and Data, Algebra, and Reasoning and Proof.

A search of the literature found no numeracy test for pre-service teachers that supplied a *profile* of ability across all strands of the mathematics curriculum. Our work not only provides summative assessment but also details supplementary diagnostic information: what errors pre-service teachers make, what strategies and misconceptions can be inferred from their errors, what the ability levels of students who hold the targeted misconceptions are, and how the errors can be used so that the pre-service teachers can re-organise their mathematical understanding.

Format and Test Item Construction

Written tests can be presented in multiple-choice format or open-ended format or a combination of both. At the heart of format decisions lies a consideration of the qualitative difference between the *selection* and *supply* of a response by an examinee. In multiple-choice tests, the responses are limited to the selection provided by the test writers. In open-ended tests, the responses supplied are theoretically not limited. Qualitatively, the latter format is richer because the examinee's response is not prompted or suggested by a list of possibilities. On the other hand, the limited number of responses available in multiple-choice formats is more manageable in terms of analysis. It is the choice of distracters that is paramount in making the multiple-choice format reliable and useful diagnostically as a first-line assessment instrument.

A range of mathematics education research on children's and teachers' knowledge and errors informed the writing of test items and distracters (e.g., Ashlock, 2002; Coben, 2003; Hart, 1981; Ma, 1999; Rowland, Heal, Barber, & Martyn, 1998; Ryan & Williams, 2000; Williams & Ryan, 2000). Items were written with diagnostic coding for most distracters; distracters included known errors, diagnostic misconceptions, and incorrect strategies. It was also seen to be important to provide adult contexts for test items and to take advantage of the higher reading ability of adult students.

Substantive and Syntactic Knowledge

It is mandatory in England for teacher education courses to provide an *audit* of students' mathematical knowledge during their training. Some key research on the subject matter knowledge (SMK) of primary pre-service teachers in England (e.g., Goulding, Rowland, & Barber, 2002; Rowland, Martyn, Barber, & Heal, 2001) informed item development in terms of tapping connected SMK including knowledge of (substantive) and knowledge about (syntactic) mathematics.

Some states in the United States of America (USA) use professional assessments for beginning teachers as part of their teacher licensure process (Gitomer, Latham, & Ziomek, 1999). The widely-used PRAXIS (2003) pre-professional skills tests (PPST) for mathematics also informed test development.

The emphasis in the *PPST: Mathematics* test is on interpretation rather than computation (p. 58). Its questions are chosen from five categories: conceptual knowledge, procedural knowledge, representations of quantitative information, measurement and informal geometry, and formal mathematical reasoning in a quantitative context.

The multiple-choice items for *TEMT* were written to test both substantive and syntactic knowledge of the constructed primary teacher curriculum. The items developed for the test are not dissimilar from items used for similar purposes in England and the USA and include both mathematical calculation and interpretation.

Method

The test items were trialled in an Australian university with a large cohort of pre-service primary teachers (N = 426). The writing team met initially with the mathematics lecturer with a test specification and a preliminary item bank document for discussion. The proposed items involved a more comprehensive coverage of the mathematics curriculum than the university's *Basic Skills Test* had covered previously (where Number had been the focus). All items were then scrutinised by an experienced team of item writers and they were then sent to the university lecturer for comment. Only minor amendments were suggested for a few items. The items were deemed thus to have face validity.

Using a bank of 105 items, three equivalent forms of the test (A, B and C) were constructed with 15 link items (L) across the forms. Each test form contained 45 multiple-choice items. The tests were timed for a 45-minute testing period and the use of a calculator was not allowed. Computational demand was at *CSF* level 6 (England GCSE equivalent) and other items asked for selection of the appropriate calculation. The items were weighted across the curriculum strands: Number (16 items in each test), Measurement (8), Space and Shape (8), Chance and Data (6), Algebra (5), and Reasoning and Proof (2). The strands were 'randomised' throughout the test and most link items fell within the first half of the test and all before the thirtieth item. Marks were not deducted for incorrect responses.

Students across three different undergraduate and graduate courses took a test form in the first few weeks of the first year of their teacher education degree at a university in Australia in 2004. A second cohort of 86 pre-service teachers in the second year of their undergraduate course in England took a *TEMT* test form in 2005. They were then given their personalised response profile to use to support their subject knowledge development.

Test Analysis

A Rasch analysis (Rasch, 1980; Wright & Stone, 1979) was undertaken using *Quest* (Adams & Khoo, 1996). The Rasch model is an item-response theory model that "can help transform raw data ... into abstract, equal-interval scales. Equality of intervals is achieved through log transformations of raw data odds, and abstraction is accomplished through probabilistic equations" (Bond & Fox, 2001, p. 7). The Rasch simple logistic dichotomous model was used; in the scaling of the items we were interested only in correct/incorrect response. *Quest* provides item estimates (item difficulty estimates with the mean difficulty set at zero), case estimates (student ability estimates), and fit statistics.

Test Reliability, Goodness of Fit, and Test Equating

The Australian sample size was 426 (cases) and the TEMT scale contained 105 questions (items). Any item omitted by pre-service teachers (students) on their particular test form was treated as incorrect rather than missing since most students completed their 45-item test. A fatigue factor was not evident. The estimates were found for all items on the scale (see Table 1 summary). All items were thus calibrated in terms of difficulty. There was high internal consistency; that is, 97 percent of the observed estimate variance is considered true. There were no items with zero scores and no items with perfect scores.

Table 1
Summary of Item Estimate Statistics

Mean	0.00
SD	1.47
SD (adjusted)	1.45
Reliability of estimate	0.97

The means and standard deviations of the infit (weighted) and outfit (unweighted) statistics in their mean square and transformed (t) forms are shown in Table 2. When the expected value of the mean squares is approximately 1 and the expected value of the t-values is approximately zero, the data are compatible with the Rasch model. Thus, the data are shown to be compatible with the model.

Table 2
Summary of Item Fit Statistics

	Infit mean square	Outfit mean square	Infit-t	Outfit-t
Mean	1.00	1.10	-0.02	0.16
SD	0.12	0.51	1.47	1.28

The estimates were found for all cases on the scale (see Table 3). The ability of each student is thus calibrated. There was high internal consistency; that is, 88 percent of the observed estimate variance is considered true. There were no cases with zero scores and no cases with perfect scores.

Table 3
Summary of Case Estimates Statistics

Mean	0.64
SD	1.14
SD (adjusted)	1.07
Reliability of estimate	0.88

The means and standard deviations of the infit (weighted) and outfit (unweighted) statistics in their mean square and transformed (t) forms are shown in Table 4. The data are shown to be compatible with the Rasch model.

Table 4
Summary of Case Fit Statistics

	Infit mean square	Outfit mean square	Infit-t	Outfit-t
Mean	0.99	1.07	-0.01	0.15
SD	0.16	0.84	0.92	0.94

The *item map* (see Figure 1) provides a logit scale on which both items and cases are calibrated. The distribution of case estimates (student ability) is shown on the left hand side of the map.

```
TEMT Person Ability Estimates                      TEMT Item Difficulty Estimates
-----------------------------------------------------------------------------------
 5.0                                 |
                                     |
                                  X  |
                                  X  |
                                     |
 4.0                                 |
                                  X  |
                                     |
                                     |
                                XXX  |
                                     |  A23S    C30R
 3.0                             XX  |
                                 XX  |
                                 XX  |  B35M    C38S
                                XXX  |  B37S    C40C
                           XXXXXXXX  |  A27C    C23S    C36M
 2.0                          XXXXX  |  L20M    C37S
                              XXXXX  |
                          XXXXXXXXX  |  A25S    A39C    A42N
                     XXXXXXXXXXXXXX  |  A41A    A44N    B23S    B30R    B38S
                         XXXXXXXXXX  |  L8S     A32N
                             XXXXXX  |  L19N    A45N    B21M    B41A
 1.0               XXXXXXXXXXXXXXXX  |  L13A    B44N
                     XXXXXXXXXXXXXX  |  A40C    B9S     B24S    B25S    B45N
                    XXXXXXXXXXXXXXX  |  A33N    A38S    B43N    C5M
                          XXXXXXXXX  |  A21M    A30R    A35M    C14A    C25S    C39C
                   XXXXXXXXXXXXXXXX  |  L28A    A29A    A37S    B34N    C29A    C44N
 0.0                   XXXXXXXXXXXX  |  L10S    B5M     C32N
                        XXXXXXXXXXX  |  B33N    B40C    C34N    C41A    C45N
                         XXXXXXXXXX  |  L6M     L7M     L11C    L26C    A5M     B2N
                         XXXXXXXXXX  |  B12C    B22M    C16N    C43N
                         XXXXXXXXXX  |  L4N     L15R    B16N    B39C
                             XXXXXX  |  L3N     A2N     A16N    A24S    A34N    B42N
-1.0                         XXXXXX  |  L17N    L18N    A9S     A43N    B31N    C42N
                                XXX  |  A1N     A22M    C12C
                                XXX  |  B36M    C31N
                                XXX  |  A14A    A31N    A36M    B32N    C1N
                                     |  C35M
-2.0                              X  |  B29A    C22M
                                  X  |  A12C
                                     |  B14A    C21M    C24S
                                  X  |
                                     |  C27C
                                     |
-3.0                                 |
                                     |
                                     |  B27C
                                  X  |  C33N
                                     |
-4.0                                 |
                                     |
                                     |
                                     |
                                     |
                                     |  B1N
-5.0                                 |
-----------------------------------------------------------------------------------
(Not all items could be fitted to the display. Each X here represents 2 students)
```

Figure 1. Item-person map for the TEMT analysis (Quest output).

A student with an ability estimate of, say, 1.0 is likely (probability level of 0.5) to have correctly answered all items having a difficulty at the same estimate (here, 1.0). For items below/above this difficulty, the student is (progressively) more/less likely to have correctly answered the items. The distribution of item estimates (difficulty) is shown on the right hand side of the map. For example, item L13A shaded (Link item, Q13 on all forms, Algebra strand) has a difficulty estimate of 1 on the logit scale. The three test forms (A, B and C) were found to be well-equated; that is, the three forms gave the same score for the same ability (logit). This outcome was not essential as we were seeking to simply scale items. However, we did attempt to distribute the items so that the three groups of pre-service teachers experienced a test of overall equivalent difficulty.

Errors and Misconceptions

Error responses were analysed for all test items. Altogether 44 percent of the 105 items contained at least one distracter that, from the literature, is believed to diagnose a significant misconception. Ninety-three percent of these errors occurred significantly more than would be expected from students' guessing and hence provides evidence of the targeted misconception. We now discuss examples of errors made by the pre-service teachers relating to decimal place value, measure of line segments, probability intuitions, and algebraic expressions. We report both frequency and mean ability logit to cater for different audiences.

Several *Number* items targeted place value understanding: one such item was '300.62 ÷ 100'. Almost one-third of pre-service teachers gave incorrect responses (see Table 5). The errors may diagnose a misconception related to 'integer-decimal' separation (response E) where only the 'whole number' is divided or application of a faulty rule for moving decimal places (response A: 30062 and response B: 30.062). The mean ability of students making each error here indicates that pre-service teachers of a higher ability are making the 'separation' error.

Table 5
Item Analysis: '300.62 ÷ 100'

Response	Inferred Misconception/Strategy	Frequency N = 426	Mean Ability (logit)
A. 30062	'Move' decimal point/digits in wrong direction	0%	
B. 30.062	'Move' decimal point/digit incorrect number of places	6%	-0.13
C. 30.62	'Cancel' a zero	3%	-0.18
D. 3.0062	CORRECT	69%	0.98
E. 3.62	Integer-decimal separation or 'cancel 2 zeros'	22%	0.10

This separation strategy was also evident in other test items where the operation was multiplication and also where the number was mixed (integer and fraction). The misconception underlying the strategy is important because it also appears to be at the root of the well-documented 'decimal point ignored' error (Assessment of Performance Unit (APU), 1982) that children are known to make. It is this error that the pre-service teachers would be expected to target in their own teaching of school children.

A *Measurement* item uncovered fundamental misconceptions related to the measure of line segments (see Figure 2). Nearly two-thirds of the pre-service teachers gave incorrect responses to the item (see Table 6) which was similar to one given to 12, 13 and 14 year olds (Hart, 1981) with similar results.

A 6-sided figure is drawn on a centimetre square grid as shown.

The distance around the edge of the figure is:

A. 12 cm

B. More than 12 cm

C. Less than 12 cm

D. You cannot tell

Grid unit

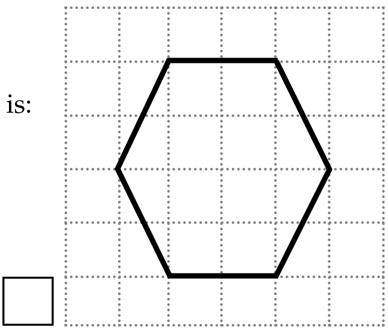

Figure 2. Length of a diagonal line segment item.

The responses here from the pre-service teachers suggest important misconceptions are at play in terms of shape, space and measurement. Only 34 percent of the sample was correct; a further 36 percent did not distinguish between the horizontal and sloping side measures of the hexagon.

Table 6
Item Analysis: Length of a Diagonal Line Segment

Response	Inferred Misconception/Strategy	Frequency N = 426	Mean Ability (logit)
A. 12 cm	Diagonal same length as side of rectangle OR *regular* hexagon prototype OR 'count squares' segment passes through	36%	0.31
B. More than 12 cm	CORRECT	34%	1.27

C. Less than 12 cm	Diagonal is smaller than side of rectangle	22%	0.24
D. You cannot tell	Visual rather than measure perception	5%	0.69
Omitted		2%	-0.14

A *Chance and Data* item on the numerical likelihood of an event again shows that pre-service teachers are making the same error as children (see Figure 3).

Here is a spinner for a game. What is the probability of it landing on a green?

A. $\frac{1}{4}$

B. $\frac{1}{2}$

C. $\frac{1}{3}$

D. $\frac{2}{3}$

E. $\frac{3}{4}$

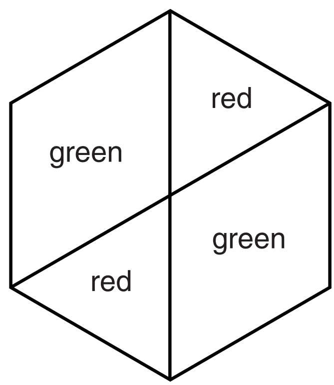

Figure 3. Probability intuitions item.

This item uncovers the equi-probability intuition that if there are two outcomes (in this case, red or green) they are assumed to be equally likely to occur (Green, 1982; Lecoutre, 1992). Table 7 shows 17 percent of the pre-service teachers making the equi-probability error.

Table 7
Item Analysis: Probability Intuitions

Response	Inferred Misconception/Strategy	Frequency N = 426	Mean Ability (logit)
A.	Unknown or Random response	3%	-0.53
B.	Equi-probability intuition	17%	0.12
C.	Unknown or Random response	7%	0.28
D.	CORRECT	69%	0.93
E.	Incorrect estimate of ‘green’ area	4%	-0.52

An *Algebra* item 'multiply n + 4 by 5' showed that only 56 percent treated n + 4 as a closed object where both letter and number are multiplied by 5 (see Table 8).

Table 8
Item Analysis: Algebraic Expressions

Response	Inferred Misconception/Strategy	Frequency N = 426	Mean Ability (logit)
A. 20	Variable ignored	2%	-0.49
B. 5n	Unknown or Random response	1%	-0.68
C. n + 20	n + 4 not seen as closed object: Letter not used	24%	-0.30
D. 5n + 4	n + 4 not seen as closed object: Multiply variable only	16%	0.14
E. 5n + 20	CORRECT	56%	1.21

This item is similar to one given to 12, 13 and 14 year olds (Hart, 1981) with similar results. The two interesting errors are 5n + 4 made by 16 percent and n + 20 made by 24 percent of the pre-service teachers. The higher student mean ability suggests that 5n + 4 is a more sophisticated error.

We have discussed in detail above examples of errors in different strands of mathematics. We also found that some pre-service teachers have place value misconceptions, have bugs in whole number and decimal computation (e.g., subtract smaller-from-larger digit, have problems with zero), do not recognise that fraction parts must be equal, have bugs in fraction computation (e.g., add or subtract numerators and denominators), misinterpret calculator displays (e.g., the 'remainder'), find reverse computation problems difficult (e.g., when finding a missing percentage), use an additive strategy for ratio, have scale misconceptions (e.g., count the 'tick marks' rather than the 'gaps'), use scale prototypes (e.g., treat all scales as unitary), use incorrect conversions (e.g., 100 minutes in an hour, 100 grams in a kilogram), misinterpret data tables, have statistical misconceptions (e.g., the mean average must appear in the data set), reverse Cartesian co-ordinates, use graph prototypes (e.g., all straight line graphs are of the form y = mx) and generalise a rule on a single x-y data point in a graph or table. There were also errors in spatial and measurement vocabulary (e.g., perpendicular/diagonal/hypotenuse confusion and area/perimeter confusion). Multi-step problems were more difficult for students than single-step problems.

The errors and strategies uncovered by TEMT can, we believe, be used positively by pre-service teachers to re-organise their own mathematical understandings. We outline below how pre-service teachers in England made sense of their own error patterns. Mathematics educators involved in pre-service teaching could also use the range of errors in their student cohort as the basis for discussion of strategies and misconceptions and, importantly, investigate the representations their students draw on or could draw on for understanding. We think this will add to the pedagogical content knowledge of the pre-service teachers.

Pre-Service Teachers Using Their Errors for Learning

Quest also produces a *kidmap* (here called a mathsmap) that is an output for each individual identifying their correct *and* error response patterns. An example is shown in Figure 4 where Jennifer has an ability estimate of 0.90, a mean square infit statistic of 1.10 and a total score of 64.4 percent. The row of Xs (centre of the map) indicates the ability estimate of the student (0.90 in this case) and the dotted lines are drawn at ± 1 standard error.

The test items are plotted on the central scale at their difficulty level in logits. The items not achieved by the student are plotted on the right-hand side of the map. The actual response made for each incorrect item is indicated in parentheses: for example, Jennifer would have been expected to have achieved item 35 (below the lower dotted line) but responded incorrectly with option 5. 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 interesting. For Jennifer the errors indicated in the bottom quadrant are particularly interesting because she was expected to have responded correctly for these items, so they might indicate gaps or 'bugs' in her knowledge.

Making Sense of the Mathsmap

A cohort of 86 pre-service teachers in the second year of their undergraduate course in England was given a TEMT test. The students were then given their individual mathsmap, a list of the item descriptors (see Table 9 for Jennifer) – but not the test items – and a detailed instruction sheet on how to read their mathsmap (see brief excerpt in Figure 5).

Table 9
Analysis of Jennifer's 'Easier not Achieved' Items

Item Response	Item Difficulty	Item Description	Inferred Misconception/ Strategy
8(2)	0.95	Algebra: general statements	Variable as specific number
33(4)	0.68	Number: Identifying ratio within several ratios	Additive tendency
35(5)	0.43	Number: Calculating surface area	Area/volume confusion

M A T H S M A P

Student:	Jennifer	ability:	0.90
group:	all	fit:	1.10
scale:	all	% score:	64.4

Harder Achieved						Harder Not Achieved		
				25				
						28(3)		
						13(1)		
				42		39(1)		
						27(3)	41(2)	44(1)
						5(1)	32(5)	
				12				
				45				
					XXX	8(2)		
				40		33(4)		
			38	30				
			23	15		35(5)		
			37	29				
						6(3)		
		14	4	3		18(3)		
				7				
	22	17	2	1		9(1)		
43	4	26	19	11				
			24	10		16(2)		
				36				
				31		21(1)		
				20				
Easier Achieved						Easier Not Achieved		

Figure 4. Mathsmap for Jennifer.

6(3)	0.11	Space: Cartesian co-ordinates	Co-ordinate reversal
9(1)	-0.69	Reasoning: logic	Triangle prototype
18(3)	-0.30	Measurement: grams to kilograms	100g is 1 kg
16(2)	-1.09	Number: Fraction representation	Unequal parts of whole treated as equal
21(1)	-1.56	Algebra: words to symbols	'more than' implies multiply

The test items were withheld so that the curriculum area indicated by the descriptor was targeted for study by the student in a broad sense rather than in terms of item-specificity. The pre-service teachers were also asked to complete a short questionnaire about the mathsmap. They reported that their mathsmap was initially a puzzle but once they had read the detailed instructions it made sense.

> 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. (Charlene)

> At first I found it quite confusing but after I printed it out and was able to write on it, I was able to identify where I needed to improve. (Jill)

> I was a little confused at first but after reading the accompanying text I found it intriguing. (Matt)

The pre-service teachers were asked if the mathsmap seemed to be 'correct' for their understanding of their own knowledge and whether there were any surprises. Generally the students reported it was accurate but several reported some surprises:

> It's a straight cough, you could have got me bang to rights. Everything I was unsure of has been highlighted. (Matt)

> Mostly correct, however I would say it identified some areas of weakness and some areas where silly mistakes were made in exam conditions. (Daniel)

> I knew more than I thought I did. (Davina)

> Yes [surprised because] I thought I was slightly better at some things. I thought my percentage would have been in the sixties. I did well in my maths exam last year. A bit disappointed. (Jill)

They were also asked if they would use this kind of diagnostic feedback in their own teaching. Generally they reported that the format would need to be much simpler for primary school children:

> Yes I would like to understand the misconceptions aspect to aid my own teaching. (Matt)

> Depends upon what year group I suppose, although in a simpler form it would probably be really effective with any class. (Andrew)

> Yes [I would use it] but in a different format. I think it would be difficult at first for especially younger pupils to understand. (Shana)

There were also several open comments indicating that the mathsmap would be useful in targeting their own knowledge:

> Great idea and really helpful towards my own learning targets. (Andrew)

> It was really helpful to know exactly where I stood in terms of subject knowledge as I haven't done maths as a subject for a long time. (Davina)

Two of the pre-service teachers, Lorna and Charlene, volunteered for an interview about their responses in the two key quadrants of the mathsmap (see Figure 5). The items in the top left quadrant are the 'harder achieved' items. The correct responses here may suggest guessing in the multiple-choice test format or an unexpected area of strength. Lorna confirmed recent targeting of *Shape and Space* indicated in the top left quadrant of her mathsmap while on teaching practice because she knew already this was an area of weakness. Charlene

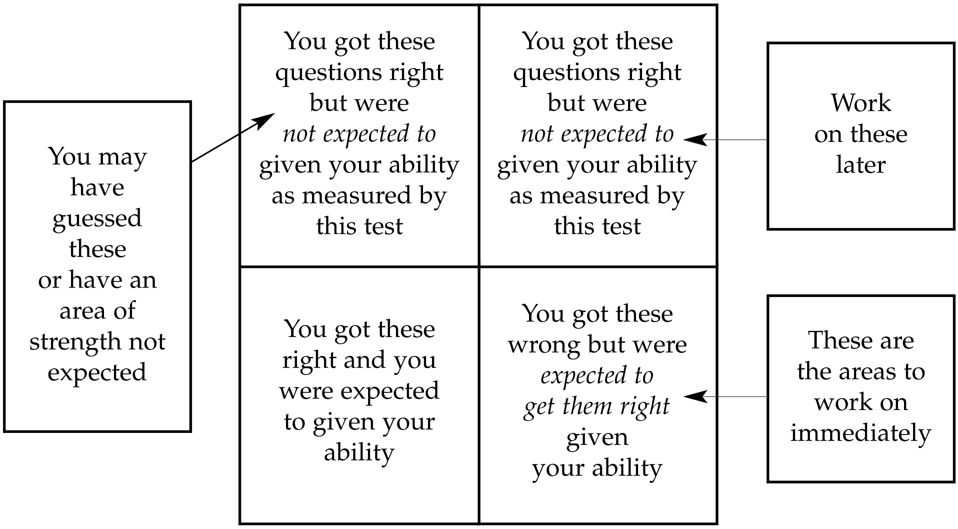

Figure 5. Summary of 'How to read your mathsmap'.

confirmed that her 'easier not achieved' items in the bottom right quadrant made sense as items she should have answered correctly:

> I mean, they looked like the sort of things that I would – probably would have had problems with or made a silly mistake on, like the decimal point. ... And also probably with that question because it's [reading from the item descriptor] 'measuring, in lengths, mm, cm and metres' so that will be converting, which is easy for me to make a mistake in. ...'Cos sometimes I try and think too advanced for the questions, 'cos I did [Advanced Maths] not very well, but I do sometimes think there's more to it than what's there. (Charlene)

Charlene knew already that some of her mistakes were a result of expecting test questions to be harder than they were. As the information was diagnostic only she was not concerned and on reflection was sometimes able to predict the mistake she had made.

Discussion

Knowledge of the common mathematical errors and misconceptions of *children* can provide teachers with an insight into children's thinking and a focus for teaching and learning (Black & Wiliam, 1998; Hart, 1981; Ryan & Williams, 2003). The errors and misconceptions made by *pre-service* teachers were used here to inform either personal development or collective treatment during pre-service teacher education. Teacher errors deserve attention not least to avoid transfer to children in schools. Errors provide opportunities for pre-service teachers to examine the basis of their own understanding so that knowledge can be re-organised and strengthened.

Errors uncovered by the ACER *TEMT* could form the basis of *pre-service* teacher group discussion; considering why the given reasoning is correct or incorrect, what warrant is presented to support a claim, and what mathematical 'tools' or artefacts are called on to demonstrate or help to re-organise understanding. This focus could be of value to a beginning teacher and to the tertiary educator seeking to gain insight into mathematical misunderstandings. A teacher educator could use cohort patterns as the basis for conflict peer group discussion of different conceptions (Bell, Swan, Onslow, Pratt, & Purdy, 1985; Ryan & Williams, 2000) to support pre-service teacher learning and to model good practice. Within group discussion, tertiary students can be asked to listen to others via discussion, justification, persuasion and finally even change of mind, so that it is the students who reorganise their own conceptions. Toulmin's (1958) model of argument is helpful here and a range of errors is valuable in such conflict discussion.

For example, the separation strategy (indicated by "300.62 ÷ 100 = 3.62" in Table 5) is suitable for such discussion, where the meaning of number and division are paramount. What representations do different tertiary students draw on to justify their claims? Which representations are successful in shifting or strengthening a conception? For a pre-service teacher, it is the use of *representations* that may shift procedural behaviour towards conceptual understanding. Representations are the life-blood of teaching and the basis of pedagogical content knowledge.

Pedagogical content knowledge is characterised as including "the most useful forms of representation of ... ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensible to others" (Shulman, 1986, p. 9). We believe that the beginning teacher needs to first make the subject comprehensible to him/herself – to examine the "veritable armamentarium of alternative forms of representation" (p. 9) so that mathematics learning is modelled dynamically as change, re-organisation and confirmation of ideas.

Conclusion

Our two-fold aim was: (i) to create a bank of valid and scaled multiple-choice items so that custom-made tests could be constructed of various difficulties and foci; and (ii) to create and validate test items that provide formative feedback, through the reporting of errors with diagnostic value. We have shown that it is possible to construct an instrument designed for the measurement of teachers' mathematics subject knowledge that also has diagnostic properties, by selecting and calibrating items that have diagnostic potential (mainly from the literature on children's misconceptions) in the test construction process. Many items revealed that significant proportions of cohorts on entry to initial teacher education have the same errors, misconceptions and incorrect strategies as children. It was further illustrated that a *mathsmap* can be used as a tool for identifying an individual pre-service teacher's profile of attainment and errors, hence providing automated feedback of potential diagnostic value to them. Work is continuing on how pre-service teachers use their mathsmap to develop subject matter knowledge and the effect this has on their pedagogical content knowledge.

Acknowledgements

We would like to thank all the people who helped to facilitate this research in Australia and England: colleagues and lecturers who commented on test items and the test frame, lecturers who gave us access to their students and conducted the tests, their students who were willing to provide their data for research purposes, students who responded to questionnaires and several who also gave hours of their own time for in-depth interviews without expectation of return. As always, without such generosity our research would not have been possible.

References

Australian Council for Educational Research (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.

Assessment of Performance Unit (1982). *A review of monitoring in mathematics 1978-1982*. London: Department of Education and Science.

Ashlock, R. B. (2002). *Error patterns in computation: Using error patterns to improve instruction* (8th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Bell, A. W., Swan, M., Onslow, B., Pratt, K., & Purdy, D. (1985). *Diagnostic teaching: Teaching for lifelong learning*. Nottingham, UK: Shell Centre for Mathematics Education.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: Department of Education and Professional Studies, King's College.

Board of Studies (1995). *Curriculum and Standards Framework: Mathematics*. Melbourne: Board of Studies.

Board of Studies (2000). *Curriculum and Standards Framework II: Mathematics* (revised edition). Melbourne: Board of Studies.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Coben, D. (2003). *Adult numeracy: Review of research and related literature. Research review*. National Research and Development Centre for Adult Literacy and Numeracy. Retrieved February 3, 2004, from http://www.nrdc.org.uk/uploads/documents/doc_2802.pdf

Curriculum Corporation (1994). *Mathematics – A curriculum profile for Australian schools*. Melbourne: Curriculum Corporation.

Department for Education and Employment (1998). *Teaching, high status, high standards: Circular 4/98*. London: Her Majesty's Stationery Office. Retrieved February 3, 2004, from http://www.dfes.gov.uk/publications/guidanceonthelaw/4_98/annexd.htm

Gitomer, D. H., Latham, A. S., & Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Educational Testing Service.

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.

Green, D. R. (1982). *Probability concepts in 11-16 year old pupils* (2nd ed.). Loughborough, UK: Centre for the Advancement of Mathematical Education in Technology.

Hart, K. (1981). *Children's understanding of mathematics 11-16.* London: John Murray.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in 'purely random' situations. *Educational Studies in Mathematics, 23*, 557-568.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.

Rowland, T., Heal, C., Barber, P., &. Martyn, S. (1998). Mind the 'gaps': Primary teacher trainees' mathematics subject knowledge. In E. Bills (Ed.), *Proceedings of the British Society for Research in Learning Mathematics day conference at Birmingham* (pp. 91-96). Coventry, UK: University of Warwick.

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). The Netherlands: University of Utrecht.

Ryan, J., & Williams, J. (2000). *Mathematical discussions with children: Exploring methods and misconceptions as a teaching strategy*. Manchester, UK: Centre for Mathematics Education, University of Manchester.

Ryan, J., & Williams, J. (2003). Charting argumentation space in conceptual locales: Tools at the boundary between research and practice. *Research in Mathematics Education, 4*, 89-111.

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

PRAXIS series, Professional Assessments for Beginning Teachers (2003). *Study guide for the pre-professional skills tests* (PPST) (2nd ed.). Princeton NJ: Educational Testing Service.

Teacher Training Agency (2003). *Qualifying to teach: Professional standards for Qualified Teacher Status*. London: Teacher Training Agency.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

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

Wright, B., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press.

Authors

Julie Ryan, Faculty of Health, Social Care and Education, Manchester Metropolitan University, 799 Didsbury Road, Manchester, M20 2RR, UK. Email: <J.T.Ryan@mmu.ac.uk>

Barry McCrae, Australian Council for Educational Research, 19 Prospect Hill Rd, Private Bag 55, Camberwell, VIC, 3124. Email: <mccrae@acer.edu.au>