A Developmental Scale of Mental Computation

Rosemary Callingham
University of Tasmania
<rosemary.callingham@utas.edu.au>

Alistair McIntosh
University of Tasmania
<alistair.mcintosh@utas.edu.au>

Research has not identified a general hierarchy of difficulty for calculations performed mentally. This study aimed to provide baseline data about students' mental computation competence across grades 3 to 10. Rasch modelling techniques were used to develop a scale of mental computation competence, based on 1452 students' responses to 238 mental computation items organised into overlapping tests. Eight levels of competence were identified. Grade based scales were also developed for grades 3 to 8. These suggested possible curriculum effects since particular items behaved differently across the grades.

The importance of mental computation skills for students has been recognised in various policy and curriculum documents for many years, and this continues to be the situation today. "Mental Computation" is one strand of the Australian Mathematics Profile (Curriculum Corporation, 1994) and mental computation is included in the National Numeracy Benchmarks (Curriculum Corporation, 2000). Surprisingly, however, in light of its recognised importance, there has been limited research into mental computation, other than basic number facts.

Efficient calculation strategies and automatic responses are essential underpinnings of mental computation (Reys & Reys, 1986). There is little information, however, about when students should be expected to have automatic recall of number facts. The Mathematics Profile (Curriculum Corporation, 1994) suggests that students should "Remember basic addition facts …" at Level 3 and "Remember basic multiplication facts (to 10 x 10)" by Level 4. These levels correspond approximately to the middle and upper grades of primary school. The Tasmanian Key Intended Numeracy Outcomes expect students at the end of grade 5 to be using strategies for multiplying numbers to 10 x 10 with the goal of these being committed to memory by the end of Grade 6. By the end of Year 8, students should be able to recall and use simple percentages, such as 10%, and their equivalents in decimals and fractions, and perform calculations such as _ of 20 (Department of Education, Community and Cultural Development, 1995). Whether it is reasonable to expect this level of competence in students in these grades is untested.

Although there is an accepted general hierarchy of difficulty for paper and pencil computations, no similar hierarchy or theoretical framework exists for calculations performed mentally. An Australian study considered the performance of students in grades 3, 5, 7, and 9 but did not develop a developmental scale of mental computation (McIntosh, Bana, & Farrell, 1995). This is one of the few studies that has considered students above primary age. Research has begun to focus on children's thinking and strategies when carrying out mental computation (e.g. Bana & Korbosky, 1995; McIntosh, de Nardi, & Swan, 1994). McIntosh and associates studied the intersection between number sense and mental computation and concluded that number sense and mental computation "…are indeed linked, especially after the age of 12" (McIntosh, Reys, Reys, Bana, & Farrell, 1997, p.54).

The study reported here aimed to collect baseline data about a developmental scale of mental computation across students in Grades 3 to 10. The value of such a scale is that it could provide valuable information to teachers about the appropriate sequence of instruction in mental computation, and lead to more effective ways to assess mental computation ability than the daily “instant recall” test.

Theoretical Model

Studies of mental computation have, until now, relied on facility rates, or percentage correct, as a measure of students’ ability. This implicitly assumes that there is a construct or variable that can be labelled as “mental computation ability” and that higher numbers of items correct implies a greater level of competence on this variable. However, it does not allow for direct comparison between groups of students unless they undertake exactly the same items under the same conditions.

Item response modelling, and specifically the Rasch model, uses the interaction between persons and items to determine the probability of success of each person on each item of a test. This provides a set of scores that describes the locations of persons and items along an underlying variable (Griffin, 1997). In the study reported here the underlying variable is mental computation competence and the positions of items on the scale reflect increasing competence.

The model (Rasch, 1980) can be represented by the equation:

$$P\{x \backslash \__{v}, \__{i}\} = \frac{e^{x(\_v - \_i)}}{1 + e^{x(\_v - \_i)}}$$

where:

P is the probability of a result (x) on an item (i) attempted by a person (v). In the case of dichotomous (right/wrong) responses, as in this study, x can take the value 0 or 1.

$\__{v}$ is the position of person v on the variable and is referred to as the ability parameter.

$\__{i}$ is the position of the item i on the variable and is referred to as the difficulty parameter.

Thus, the probability of success on an item is related to the difference between the person ability, $\__{v}$, and the item difficulty, $\__{i}$. Since ability is represented by a single parameter the model is regarded as unidimensional, that is only one dimension governs the responses of the persons to the items. Similarly, only one dimension governs the behaviour of the items with respect to the persons. The position of both persons and items can be computed on the same scale, using a single unit of measure that is repetitive and additive along the variable. This is the logit, the logarithm of the odds of success. In order to meet these requirements one further assumption is needed. Each item must be independent of all others so that each can be used as a pointer or indicator of position along the variable.

Under these conditions, it can be proved that the odds of success by a person on one item, i, but not on another item, j, is the difference between the difficulties of the two items, $\__{i} - \__{j}$, and is not dependent on the ability parameter. This is the basis for the claim that the Rasch model is sample independent (Griffin, 1997). This is also the basis for being able to develop a scale and place persons along it without every person having to attempt every item. Provided that each test of the target variable has a number of common items, these can be linked through a simultaneous process of calibration and equating. Calibration refers to the computation of the difficulty level of the items and the establishment of their accuracy as pointers along the variable. Equating refers to the process of linking different test forms onto a common scale.

These characteristics of the Rasch model made it a suitable approach to the problem of developing a scale of mental computation. It allowed a large pool of items to be used, and tests constructed appropriate to the grade level being tested. The nature of the items meant that the assumption of unidimensionality was well founded, and previous research studies indicated that there was a hierarchy of items so that the directional nature of the model could be met.

Research Questions and Methodology

The aim of the study was to collect baseline information about students' mental computation ability across the primary and secondary years of schooling. To this end, several research questions were formulated.

1. What are the features of a developmental scale of mental computation?
2. What happens to students' mental computation ability as they move through the years of schooling?
3. Are there particular kinds of mental computation question that might provide useful pointers to teachers about appropriate intervention?

To answer these questions, tests of mental computation were prepared for adjacent grades: grades 3-4, grades 5-6, grades 7-8, and grade 9-10. Each test had three forms containing a number of overlapping items, and each grade level test was also linked to the next grade through use of overlapping items. In total, 238 items were developed based on researchers' prior experience and available research.

Items were recorded and provided to schools on audiotape to ensure consistency of delivery. All items were only delivered orally – students did not see the written form of the question. The first part of each test consisted of 20 (25 for grades 9-10) *short* items – items with a five second response time. The second part contained *long* items – items with a 15 second response time. This allowed students time for working the question out. The total number of items given in any test was 40 for grades 3-4 and 50 for all other grades. A summary of the test format is shown in Table 1.

Table 1
Test Formats for Mental Computation Tests

	Gr 3-4	Gr 5-6	Gr 7-8	Gr 9-10
Short items	20	20	20	25
Long items	20	30	30	25

Altogether 1452 students in one high school (grades 7-10), one district high school (grades 3-10) and five primary schools (grades 3-6) attempted the tests. The student sample is summarised in Table 2. All items except one were attempted by at least 55 students and most items were attempted by over 100 students. In the short item part of the tests, nine items were attempted by students in every grade from grade 3 to grade 10. These items provided an anchor set for analysis of the data by grade level.

Table 2
Number of Students by Grade

Gd. 3	Gd. 4	Gd. 5	Gd. 6	Gd. 7	Gd. 8	Gd. 9	Gd. 10
247	233	224	215	161	93	139	140

Tests were organised by item topic, addition, subtraction, multiplication, division, and non-whole number items, rather than using a format with mixed operations. This followed a similar format to that used in other studies (McIntosh, Bana, & Farrell, 1995). Some items were also repeated in the short and long versions to see whether having more working time made a difference to the difficulty of the item.

Students' actual responses were entered into a spreadsheet to enable error analysis to be undertaken. For Rasch analysis, the items were scored correct or incorrect, with items presented to students but not answered regarded as incorrect. Using *Quest v2.1* software (Adams & Khoo, 1996), initial analysis undertook simultaneous calibration and equating of all responses to all test forms. Subsequent analyses provided sub-scales of short and long items, whole numbers, and fractions, decimals and percentages. Using the Quest estimates of item difficulty, clusters of items were identified along the variable. These were analysed for content and possible similarities in thinking strategies.

Using the subset of items given across all grades as an anchor file, a second analysis of short items by grade was carried out. The use of an anchor file fixes the difficulty level of these items and allows the other items to be adjusted with reference to these fixed values. In this way comparisons can be made across sub-groups of students – in this instance by grade (Griffin, 1997). The results reported here address only the full scale of mental computation and the short items' comparative grade scales.

Results

Overall Mental Computation Scale

Analysis of the item clusters, identified from the item difficulty estimates of all 238 items, suggested 8 levels of competence in mental computation. These are summarised in Table 3.

In general, in any one level the kinds of questions that occurred for a particular topic were of the same type, regardless of whether they were short or long items. For example, 2-digit by 2-digit addition items 60+80 (short) and 37+24 (long) both appeared in Level 5. It was noticeable that questions that were asked in both formats, to different groups of students, were very close in difficulty level. In many instances the longer response time led to a slightly higher difficulty. Although no information was gathered about the nature of the strategies used by students, this suggests that if students have a strategy they can apply it quickly, and providing additional response time does not decrease the difficulty of the item for questions at this level of complexity.

In broad terms the sequence of outcomes was not unexpected. Questions with more digits were, in general, more difficult than single-digit items and this might have been expected from the oral delivery format of the tests. Again as expected, subtraction and division were more difficult than their inverses, addition and multiplication. Problems involving decimals, percentages and fractions, not surprisingly, were at the upper end of the scale. An exception to this was 0.25+0.25, which was the only decimal problem to appear in the lower (easier) half of the scale. It is possible that students were treating this more as a “double” with little reference to the decimal point. “Doubles” and “one-half of…” appeared earlier than other related ideas, suggesting that students develop intuitive understandings of these concepts.

Subtractions of 2-digit from 3-digit numbers seemed to be problematic, even where there appeared to be a straightforward strategy. For example, 264 – 99 appeared at a higher difficulty level than 111 – 67, which was unexpected. Similarly, 100 – 68 appeared at Level 5 while 124 – 99 and 105 – 26 both lay in Level 6. It may be that students are not explicitly taught strategies such as rounding up or down to make a subtraction problem easier. Attempting to use a mental form of the written algorithm could explain the high difficulty of 264 – 99, for example.

Not surprisingly, the size of the multiplicands, as well as that of the product of multiplication seemed to affect the item difficulty. For example, while 12 x 20 and 25 x 6 were in Level 6, 25 x 40 and 30 x 60 appeared in Level 7 and 60 x 70 and 40 x 80 were in Level 8, among the most difficult items. In multiplication and addition problems, however, the orders of the addends or multiplicands did not appear to make a difference, 3 x 6 and 7 x 3 both appearing in the same level, for example. This may suggest that students at this level have an intuitive understanding of commutativity.

Table 3
Mental Computation Scale

Level	Content analysis	Example
1	Addition – bonds to 10 and near 10	1+6; 6+1; 4+7
2	Addition – bonds to 20 and 2-digit + 1-digit; simple doubles and multiplication by 2; zeroes	8+6; 16+8; 0+8; 4-0; double 30; 6x2
3	Subtraction bonds to 10 and 2-digit - 1-digit without regrouping; Multiplication single-digit x 10	8-3; 36-5; 7x10; 5x10
4	Table facts x3, x4, x5; Addition and subtraction of multiples of 10 (<200); Subtraction: bonds to 20; Addition of a simple common decimal; One-half of …	4x3; 6x5; 70-30; 120-50; 14-6; 0.25+0.25; one-half of 62
5	Table facts and simple related division facts; Multiplication: single-digit x 200, 2-digit doubles; Subtraction 2-digit - 2-digit without regrouping; Addition: 2-digit + 2-digit	6x9; 12÷4; 7x200; double 38; 65-35; 74-30; 25+99; 27+25
6	Division: inverse table facts; Addition & subtraction: 2 and 3-digit - 2-digit; addition of simple decimals (1 decimal place) and halves and quarters; Unit fraction of a whole number	72 ÷ 8; 58+34; 57-18; _ + _; 1.3+1.7; one–third of 12
7	Multiplication: 2-digit x 1-digit; decimals (1 decimal place) x 10; Division: divide by 0.1; Addition: simple equivalent fractions, straightforward decimals; Subtraction: 3-digit – 2-digit; Simple percentages.	24x3; 0.6 x 10; 2÷0.1; _ + ; 0.5+0.75; 111-67; 10% of 45
8	Multiplication and division by 0.5; Multiplication of multiple of 10 by a multiple of 10; Multiplication of decimals; Addition: simple unit fractions.	2 ÷ 0.5; 60 x 70 0.1 x 0.1; _ +

Comparative Grade Scales

Using eight items that were given across all grades as an anchor, separate scales were prepared for each grade. These were linked through the anchor items, allowing comparison of the item difficulties across the grades. Mean item difficulties and person abilities were calculated for each grade. Comparison of these measures allowed a judgement to be made of the match of the test to the student cohort. The results are summarised in Figure 1. There is an apparent mismatch of test difficulty with student ability in all grades. The tests in general were relatively easy for the primary, grades 3 – 6, students and relatively difficult for students in the high school, grades 7 – 10. The closest match was in grade 5 and the biggest mismatches in grades 4 and 6.

In some respects this is not surprising. Each test was designed for two adjacent grades, and the test content reflected likely curriculum emphases in those grades. In the high school years, a greater proportion of items addressing fractions, decimals and percentages was included and this probably explains the jump in average difficulty from grade 7 onwards. The average ability level did not change greatly in the high school grades, but fluctuated in the primary grades. This may be due to curriculum effects. Less emphasis tends to be placed on mental computation practice in high schools but is commonplace in primary schools through strategies such as “instant recall”. The jumps in ability levels in grades 4 and 6 may reflect increasing competence with the curriculum content addressed in this test of mental computation.

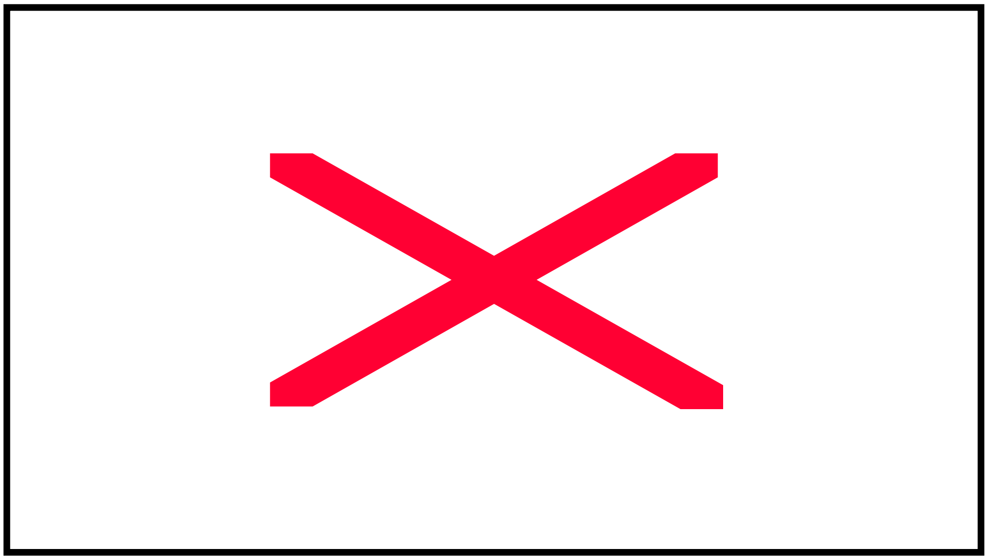

Figure 1. Mean person ability and item difficulty across grades.

This was confirmed to some extent by a consideration of items that had been given across multiple grades, although not in all grades. A summary of the difficulty levels of some of these items is provided in Table 4. The three "tables" items are notable for the relatively large drop in difficulty in grade 6 over that in grade 5, and the increase in difficulty again in the high school grades. One possible explanation is that emphasis may well be placed on these kinds of computations in the latter years of primary school. If this were so, year 6 students would be more likely to be practiced in these kinds of questions and thus experience success, lowering the relative difficulty level for these items. A relative lack of practice in the high school grades could contribute to the increased difficulty.

The behaviour of the three division items is more difficult to explain. In general, division appeared to be easier in grades 7 and 8 and rose in difficulty to some extent in grades 9 and 10. However, 30÷5 appeared to follow no pattern. It may be that division by 5 needs to be investigated further. Similarly, the final item in the table appeared to be aberrant. The difficulty level of _ + _ varied by relatively small amounts in grade 5 and the high school grades but dropped dramatically in grade 6. One possible explanation is that a fraction is on the edge of grade 5 students' understanding but is answered by grade 6 students applying a common sense approach to the worded question of "one-half plus one-quarter". By grade 7 onwards, students may have learned that fractions are "difficult", and approach fraction questions with some prejudice. Further investigation is needed into some of these unanswered conjectures.

Table 4
Item Difficulties in Logits – Anchored Across Grades

Item	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8	Gr. 9	Gr. 10
9x8			3.27	-0.20	2.55	2.46	2.43	2.38
8x7			2.36	0.48	2.86	3.11	2.70	3.83
6x9			3.15	0.83	2.78		0.17	2.26
21÷3	2.80	3.07			1.60	1.29	2.05	2.14
12÷4	1.42	1.21			0.86	0.17	1.38	0.12
30÷5	1.81	0.72			1.95		0.17	1.31
_ + _			3.09	-0.29	2.68	2.81	3.03	2.30

Discussion

Features of a Developmental Scale of Mental Computation

An eight-level hierarchy of mental computation ability was identified in this study. This progressed from simple single-digit additions to quite complex computations involving fractions and decimals. As well as single-digit additions, doubling and multiplying by two were relatively easy, first appearing in the second level of the scale. This was so even when the item was given with the largest digit as the multiplier, for example, “six times two”. Students appeared in this instance to have an intuitive grasp of commutativity. This may suggest that students were using personal strategies rather than learned responses, since other multiplication items, including multiplication by 10, appeared at a higher level. Subtraction with regrouping proved difficult, even number bonds to 20, such as 13-6, not appearing consistently in the scale until Level 4. Items such as 67-29 and 264-99 were in Level 7, along with items such as 111-67, which suggests that students did not know or could not use strategies such as rounding up to make the problem easier. Tables facts x3, x4, and x5 were consistently easier than the later ones, confirming research by Bana and Korbosky (1995). For specific multiplication facts, related division facts tended to appear in the next level, suggesting that students develop understandings of the relationship between multiplication and division relatively easily. Additional work will focus on other kinds of relationships, such as related addition and subtraction facts.

Of interest was the finding that giving additional time for students to work out answers did not seem to change the difficulty level greatly. Items repeated in the short and long sections of the test consistently appeared close together on the variable, regardless of topic. The extent to which students were working out answers very quickly as opposed to remembering answers needs to be investigated further. In some instances this seemed likely – students would not be expected to have instant recall of 0.2÷5 or _ + , for example, yet in both long and short forms these questions appeared together on the variable.

Changes in Mental Computation Ability Across Grades

Although comparative grade scales could be developed, the changes across grades are difficult to explain. In particular, the change in difficulty level of some items seems counter-intuitive, for example the sharp drop in difficulty of _ + _ in grade 6. At this stage it is hypothesised that there is some kind of curriculum effect, with students in higher grades developing learned behaviour or attitudes towards topics such as fractions, or simply becoming less skilled as a consequence of less emphasis on mental computation in high school classrooms. This might explain the behaviour of some of the tables items shown in Table 4 for example. This warrants further investigation, both through finding out about classroom factors and students' computational strategies, and will be the subject of on-going research.

Particular Items of Interest to Teachers

From a consideration of the developmental levels of mental computation identified, it seems clear that multiplication should be developed starting with doubles and "times 2", followed by multiplication by 10, the earlier tables such as x3, x4, and x5, and then later tables. The close proximity on the scale of the related division facts suggests that the multiplication and division relationship could be addressed at the same time in teaching, rather than possibly leaving division to a later grade. Similar remarks could be made about simple addition and subtraction facts.

The apparent intuitive understanding of commutativity needs also to be investigated further. Building on student's intuitive understandings of doubles, half of …, and commutativity could provide powerful ways for teachers to develop mental computation strategies.

The high difficulty level of specific items, such as 67-29, which came in Level 7, may indicate that efficient mental strategies are not being well developed, especially since this item fell in the same level as 111-67, for example, which appears to make greater cognitive demands. Similar comments could be made about some of the fractions and decimals items, and further work will address these specifically.

Conclusion

This study has provided initial evidence of a developmental scale of mental computation. Although this scale can be applied across the years of schooling from grade 3 to grade 10, there appear to be some possible curriculum effects that were not anticipated. This is shown by the behaviour of particular items across the grades. The initial scale does, however, provide some information to teachers about the order in which mental computation skills might be developed. The identification of effective mental computation strategies, and further refinement of the scale will be the subject of ongoing research.

References

Adams, R.J. & Khoo, S.T. (1996). *Quest: Interactive item analysis*. Melbourne: ACER.

Bana, J. & Korbosky, R. (1995). *Children's knowledge and understanding of basic number facts*. Perth: MASTEC, Edith Cowan University.

Curriculum Corporation. (1994). *Mathematics - a curriculum profile for Australian schools*. Melbourne: Author.

Curriculum Corporation. (2000). *Numeracy benchmarks: Years 3, 5 & 7*. Melbourne: Author.

Department of Education, Community and Cultural Development (DECCD). (1995). *Key intended numeracy outcomes*. Hobart: Author.

Griffin, P. (1997). *Using the Rasch model*. Melbourne: Assessment Research Centre.

McIntosh A., Reys, B., Reys, R., Bana, J. & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries.* Perth: MASTEC, Edith Cowan University.

McIntosh, A., Bana, J., & Farrell, B. (1995). *Mental computation in school mathematics: Preference, attitude and performance of students in Years 3, 5, 7 and 9*. Perth: MASTEC, Edith Cowan University.

McIntosh, A., de Nardi, E. & Swan, P. (1994). *Think mathematically!* Melbourne: Longman Cheshire.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (expanded edition). Chicago: The University of Chicago Press. (Original work published 1960)

Reys, B.J. & Reys, R.E. (1986). One point of view: Mental computation and computational estimation – Their time has come. *Arithmetic Teacher,* 33(7), p. 4-5.