University of Strathclyde

Department of Psychology

Name: Roy Hunter

Class: Individual Differences

Title: Design and pilot study of a mental arithmetic attainment test

Group: Deborah Gray, Frances McMillan, Marian Shah

Due date: 30th March 2005

I affirm that this report is my own work, and does not include any unacknowledged material taken from another source.

Signed: Date: 28 September
2006

Mental arithmetic is an important skill. It is preferable to be able to work out sums in your head without the aid of a calculator if, for instance, you are a scuba diver calculating your remaining air reserves and your decompression times, or an emergency carpenter working out the maximum spacing of fixings when securing a flapping roof or a broken window in a high wind.

Even more mundane activities use mental arithmetic, like shopping: calculating the relative prices of different sized bottles and cans of food; working out the fat, salt and sugar content of food from the often cryptic information given on the packaging; and making sure you have enough money to pay for it all.

In terms of theories of multiple, or divided, intelligences, even conflicting theories like Thurstone's (1938) and Spearman's (1904) both allow the distinctness of mathematical abilities and attainments.

All children are taught arithmetic at school. They are taught to add and subtract, and they are taught their multiplication tables, from 1 x 1 up to 12 x 12. Using these basic building blocks, they can carry out quite complex multiplications according to the formal rules they are taught.

Children are taught arithmetic in an order of difficulty that relates to the number of procedures they have to perform in order to arrive at the answer to a problem[1] (Miller, Craig & Cocking, 1995 – 97).

To solve 13 x 3 takes three procedures - two multiplication table tasks and one addition task:

1.) 3 x 3 = 9
2.) 3 x 1(0) = 3(0) [brackets indicate numbers not being manipulated]
3.) 9 + 30 = 39

---

[1] Although this approach is widely used throughout the education system, and has been mentioned in psychological literature, no one author stakes a claim to have invented the system. The authors of this report are familiar with the system from marking their children's homework.

To solve 14 x 3 takes four procedures: two multiplication table tasks and two addition tasks:

1.) 3 x 4 = 12 [this is written down as 2 with the 1 carried over]

2.) 3 x 1(0) = 3(0)

3.) 2 + 30 =32

4.) 3(2) + 1(0) = 4(2)

To multiply 27 x 19 takes six procedures:

1.) 7 x 9 = 63 [3 with 6 carried over]

2.) 2(0) x 9 = 18(0)

3.) 180 + 3 = 183

4.) 18(3) + 60 = 24(3)

4.) 1(0) x 27 = 27(0)

5.) (2)43 + (2)70 = 113 [13 with 1 carried over]

6.) 2(43) + 2(70) + 1(00) = 513

As the child's attainment and confidence progresses, they are taught harder and harder sums with more and more stages to them, until they understand the systematic approach well enough to tackle any sum that is presented to them.

There are other ways to carry out these tasks, however. Many people will use a stratagem of their own instead of using the formal rules. For instance, to multiply 27 x 19=, they might instead calculate it as (27 x 20) – (27 x 1) = 513. The answer is the same, but the method, although the sum looks more complicated, is much simpler:

1.) 27 x 2(0) = 54(0)

2.) 540 – 27 = 513

Because all people who have completed primary education and learned their multiplication tables have acquired this formal system for written arithmetic, and in addition our own strategies for certain calculations, it should be possible to use the teaching system to test individuals' attainment in mental arithmetic tasks by using it in the design of a psychometric test.

The limited capacity of working memory (Miller, 1956) means that only a finite number of stages can be calculated before capacity of working memory is exhausted, and the person either gets it wrong or declines to attempt an answer. By measuring the number of stages of multiplication people can manage in their head, we hope to evaluate their ability to carry out complex arithmetical tasks.

The test will use two different methods of presenting multiplication sums: using numbers and using words. Because complex arithmetic problems are only taught in schools by using numbers and not words, it is anticipated that more people should use strategies to solve word-presented multiplications than numerically-presented multiplications. The pilot study, however, will tell us whether this is indeed so. Ultimately, though, the designers must bear in mind Edwin Boring's ( )warning that 'intelligence is what intelligence tests measure':

If the pilot study is successful, and the method does produce a viable means of testing mental arithmetic attainment, it should be possible to encode the method into a computer web-based application, which will find the level of difficulty up to which a person can perform by progressing up the levels until they start to get problems consistently wrong, then going back down a level of difficulty to see if they are still capable at that level.

By using a greater number of stimulus problems, the reliability of the method should be greater, and it would be more accessible to a greater number of people (although it would be difficult to check that they did not use a pencil and paper, or a calculator!).

Method

Test Design

This pilot study will use a small number but a wide difficulty range of multiplication problems, from three formal procedures up to ten formal procedures (see appendix A). Working on the premise of Miller's (1956) 'magical number' for working memory: 7 plus or minus 2; more or less everyone should be able to perform a three-task problem, and almost no-one should be able to perform a ten-procedure problem.

In the formulation of these problems, multiples of five and ten were avoided, as they were judged too easy to work out by using strategies, as were numbers ending in one. All the problems include at least one number greater than twelve, so that even with a good recall of multiplication tables, there are additional tasks to be undertaken in order to complete the problem.

The test uses a repeated-measures design, where the individual is given an opportunity to attempt every level of stimulus problem using both methods of presentation, starting from the easiest problems and working up to the most difficult problems. There is no time limit for each attempt: the limited time span of working memory should impose its own limits on individuals' performance.

Initially, the pilot test will use two sets of multiplication problems: 'A' problems and 'B' problems; as stimuli, and they will each be presented in two different manners: as words; or as numbers. By presenting some individuals with 'A' problems as words and 'B' problems as numbers; and some individuals with 'A' problems as numbers and 'B' problems as words, any unplanned differences in difficulty between the two sets of problems should be made more visible.

Once the test has been more fully developed, a greater number of stimulus problems with the same range of difficulty would be used, so that the test could be administered more than once to the same person without the possibility of the answers to the same questions being recalled by participants. This would make it suitable as a 'before and after' test for psychological and educational research, as well as a straightforward psychometric test.

Responses are coded as correct, incorrect, or no attempt made, although only correct scores are counted for the purposes of scoring, and the datum point for scoring the individual's attainment is the mean score from the pilot test. This point generates a score of one hundred, with each standard deviation above or below the mean score from the pilot being plus or minus ten points, in a similar scoring technique to the derivation IQ test (Cooper, 2002, p.186-7).

Pilot Study Design

The pilot study is a repeated-measures study, using the different difficulty levels of problems, and the different methods of presentation as independent variables, and overall scores as the dependent variable. The data collected is discrete and on a ratio scale.

Pilot Study Participants

Sixteen participants aged between 17 and 67 took part in the study. Nine were women and seven were men. Their educational attainments were mixed, but they had all passed through the British school education system, and had therefore been taught arithmetic using multiplication tables and the formal methods described in the introduction.

Pilot Study Materials

An information sheet describing the experiment was given to participants at the outset. Problems were presented on flash cards, eight written word problems and eight number problems. There were two separate sets of cards (see appendix B), reflecting the two sets of problems and the two methods of presentation described above. The experimenter had a question-and-answer sheet (see appendix C). Participants were not allowed to have any sort of materials like paper or pencils to assist their calculations.

Pilot Study Method

Having secured their consent for the test, participants were given the information sheet which described the experiment, and also told them that they were not expected to get all of the problems correct. Participants were tested alone, and were shown the stimulus problems starting from the 3-procedure problems and working towards the 10-procedure problems.

The experimenter recorded their response as correct, incorrect or no attempt made (the distinction between incorrect or no attempt made is not significant for this study, but may provide useful data to other studies), and once the participant had failed to answer two consecutive problems correctly, they were judged to have reached the limit of their attainment.

Pilot Study Results

The mean scores for 3-procedure and 4-procedure problems were both 16, standard deviation 0. The mean score for 5-procedure problems was 15, standard deviation 1.41. The mean score for 6-procedure scores was 12.50, standard deviation 0.71. For 7-procedure problems, the mean score was 9.50, standard deviation 0.71, and for 8-procedure problems the mean score was 5.00, standard deviation 1.41. The mean scores for both 9-procedure and 10-procedure problems was 0, standard deviation 0. The data is illustrated in figure 1.

The mean score for numerically-presented problems was 11.0, standard deviation 4.16. The mean score for word-presented problems was lower at 10.0, standard deviation 4.56. The data is illustrated in figure 2.

Figure 1: Means and standard deviations fo levels of difficulty

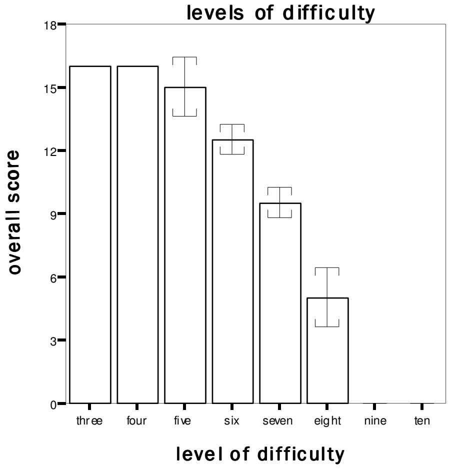


Figure 2: Means and standard deviations for type of presentation

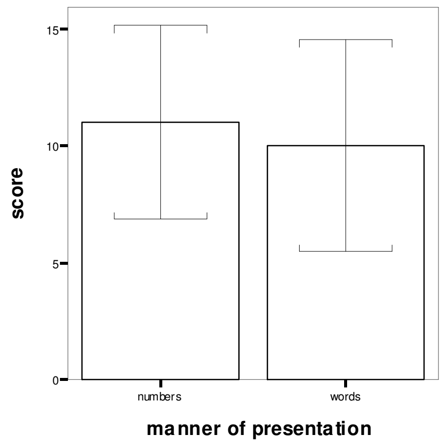
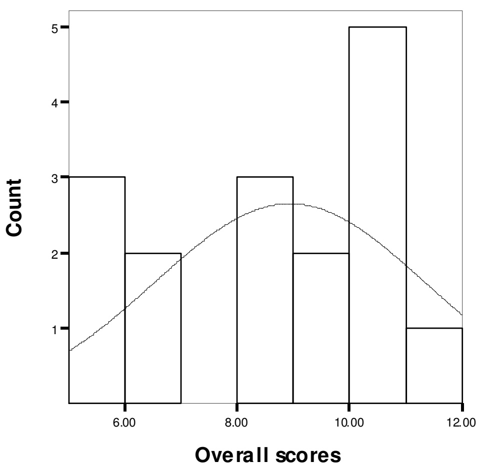


Figure 3: Distribution of pilot test scores



The overall mean score, for all levels of difficulty and both methods of presentation, was 10.5, standard deviation 4.07.

The distribution of scores from the pilot study is illustrated in figure 3.

A two-way analysis of variance was performed on the results for levels of difficulty and method of presentation (see appendix B) which returned a result of 162.00 with 7 degrees of freedom for levels of difficulty, which is significant at the 1% level [$F(7, 7) = 162.00$, $p < .005$]. The result for manner of presentation was 1.75 with 1 degree of freedom, which is non-significant [$F(7, 1) = 1.75$, $p > .05$]. A summary table for the ANOVA is shown in table 1.

Table 1: Summary data for ANOVA

	SS	df	MS	F	p
Level	648	7	92.58	162.00	< .005
Presentation	1	1	1	1.75	> .05
Error	4	7	0.571		

A *post-hoc* Tukey HSD test was carried out on the results, which showed that there were significant differences between 7-procedure and 8-procedure problems ($p < .01$), and 8-procedure and 9-procedure problems ($p < .01$), but that no other significant differences existed between consecutive levels of difficulty ($p > .05$).

Pilot Study Discussion

The results of the pilot study indicate that the number of procedures required to perform a calculation can be a good indication of how difficult it is to perform a mental multiplication problem: within the range of five to nine procedures, there were significant differences in performance. This range of difficulty could tentatively be identified as being the range between the floor and ceiling of mental arithmetic attainment.

The results also indicate, however, that there was no significant difference between the different presentations of the problems. This means that the test could be administered using either (or both) means of presentation, but that the means of presentation will not tell us anything about the strategy used to solve the problems.

There were, however, problems with the pilot study. The participant sample was very small, limiting the amount of data available to build a 'benchmark' for the psychometric test, and the distribution of the sample results is less than ideal.

The stimulus questions were formulated on the basis of their level of difficulty using formal methods to solve them, not by using individuals' informal strategies (which would in themselves have to be a subject of investigation). One of formally the most difficult questions in this study; 87 x 99 = (a 10-procedure problem); is in fact quite easily solved: (87 x 100) – (87 x 1) => 8700 – 87 = 8613.

A further limitation was placed on the pilot quite accidentally, in order to spare the participants' blushes, whereby the test would be concluded after two incorrect or refused answers: this confounded the efforts many participants who never got as far as the 10-procedure problems and may well have produced a correct answer using a simple strategy.

The small number of arithmetical problems used tended to limit the scope of the means of presentation aspect of the study: only two problems of each difficulty level were presented to each participant, which is hardly a useful measure of the difference between the two methods of presentation.

Discussion

On the basis of this first pilot study, which is by its designers' own admission flawed, it may still be possible to build a meaningful test. Using a larger sample of participants and a greater number of questions, it should be possible to construct a more focused pilot test based on the range of arithmetical problems between five and nine procedures, and ignoring the different means of presentation as a means of testing mental arithmetic attainment *per se.*

The method of scoring the test should be quite successful: the overall mean score was 10.5, (s.d. 4.07), which corresponds to the point in between the 7-procedure and 8-procedure problems. If the test uses a similar technique to scoring derivation IQ tests; the datum point of a score of 100 being the mean of the pilot study and 10 points up or down being one standard deviation away from that mean; then the majority of participants would have fallen into the plus or minus 10 points bracket.

A mean pilot score of 6.5 (which would correspond to solving 5-procedure or 6-procedure problems) would give a test score of approximately 90, and a mean pilot score of 14.5 (which would correspond to solving 9-procedure or 10-procedure problems) would give a test score of approximately 110, although it should be noted that a larger pilot study would hopefully produce a better distributed result which has implications for the scoring of this test.

This study was an attempt to design and pilot study a psychometric test to measure mental arithmetic attainment. Although some flaws emerged in both the design and execution of the pilot test, there is the potential for a useful test to be devised using the basic framework explored here. Further piloting, using a larger sample of participants and a larger number of narrower-scope questions, may produce a good basis for a successful psychometric test.

Bibliography

Boring, E. (1923), 'Intelligence as the tests test it', *The New Republic* (Washington: The New Republic)

Cooper, C. (2002), *Individual Differences,* 2nd edition (London: Arnold)

Miller, Craig & Cocking (1995 – 7), *Heinemann Mathematics Textbook,* nos. 1 – 7 (Oxford: Heinemann Educational)

Miller, G. (1956), 'The magical number seven plus or minus two: Some limits in our capacity for processing information', *Psychological Review* (63: 81 – 97)

Spearman, C. (1904), 'General Intelligence objectively determined and measured', *American Journal of Psychology* (15: 201 – 93)

Thurstone, L. (1938), *Primary Mental Abilities* (Chicago: University of Chicago Press)

Appendix A: Stimulus problems

3-task problems:	A: 13 x 3 = 39 B: 22 x 4 = 88
4-task problems:	A: 14 x 3 = 42 B: 23 x 4 = 92
5-task problems:	A: 37 x 3 = 111 B: 48 x 3 = 144
6-task problems:	A: 22 x 12 = 264 B: 23 x 12 = 276
7-task problems:	A: 23 x 13 = 299 B: 24 x 14 = 336
8-task problems:	A: 37 x 23 = 851 B: 38 x 24 = 912
9-task problems:	A: 89 x 37 = 3293 B: 89 x 47 = 4183
10-task problems:	A: 97 x 89 = 8633 B: 99 x 87 = 8613