Bills, L. (Ed.) Proceedings of the British Society for Research into Learning Mathematics 19(1) February 1999

WHAT DO MATHEMATICS TESTS TEST?

Caroline Hilton and Tim Rowland

Institute of Education, University of London

Drawing on research that set out, initially, to match Year 7 children with moderate learning difficulties with Year 1 mainstream children, we consider some difficulties with the use of standardised tests. Information gleaned from clinical interviews with these children lead us to question the value of raw test scores in assessing mathematical skills and understanding.

INTRODUCTION

According to Harry Daniels, many teachers faced with teaching low-attaining pupils with learning difficulties focus primarily on 'narrowly defined outcomes at the expense of contexts for learning' (Daniels, 1993, p.66). This view is supported by Derek Haylock (1991) who suggests that low-attaining pupils spend too much time learning routines and procedures at the expense of the development of mathematical understanding. This strategy, Daniels suggests, is due to lack of confidence and competence on the part of teachers and schools when planning a mathematics curriculum for children with learning difficulties.

The purpose of this study was to compare the performance and strategies of a group of Year 7 children with moderate learning difficulties (MLD) with a matched group of Year 1 children in a mainstream primary school. Ten children, drawn from two schools in the same North London borough, took part in the study. Five of the children formed a complete year 7 class in a school for MLD children. The other five children were selected from a Year 1 class in a primary school. All the primary school children were taken from one class in order to limit threats to internal validity. The children were considered in pairs, one from each group, matched by sex and test score. The number of participants in the study was determined by the fact that the Year 7 class contained only five children.

With the aim of gaining some understanding of the nature of their mathematical thinking, each child was then interviewed using the clinical method. Analysis of the interview transcripts revealed evidence which suggests that assessment of children's mathematical skills and understanding requires more in-depth analysis than a traditional standardised test can hope to provide.

MATERIALS

The tests used were those selected by the MLD school for the purpose of monitoring children's attainment and progress in number. These are the *Basic Number Diagnostic Test* (Gillham, 1996) and the *Basic Number Screening Test* (Gillham and Hesse, 1976). These consist of items to be administered on a one-to-one basis by a competent tester – one of the researchers in this instance. The test items to be completed by the children were enlarged in order make them more accessible.

PROCEDURE

All five children from the MLD group and twenty-four from the primary class were first tested individually by the researcher using the *Basic Number Diagnostic Test*. Those at or near ceiling on this test were then given the *Basic Number Screening Test*. Field notes were taken during the testing phase of strategies employed by the children in responding to test questions. On the basis of the test results, five children from the primary school were selected to match as closely as possible the five MLD children by test score and by gender (Table 1).

Primary Year 1				MLD Year 7			
Student	Sex	Chronological age	Test Score	Student	Sex	Chronological age	Test Score
P1	M	5:7	13	M1	M	11:8	9
P2	M	5:8	27	M2	M	11:10	26
P3	M	6:2	40	M3	M	11:7	42
P4	F	6:3	46	M4	F	11:10	47
P5	F	6:3	48	M5	F	11:11	50

Table 1: Matching of students by gender and test score

CLINICAL INTERVIEWS

The clinical interview method was chosen to investigate in some depth the quality of mathematical thinking of the ten children. The clinical method is appropriate for the purposes of identifying (eliciting), describing and accounting for cognitive processes (for details see Chapter 1 of Ginsburg *et al.*, 1983): Whereas standard mathematical assessments are primarily designed to monitor students' progress and standards within schools, the clinical interview recognises the "role of language and the importance of clarification of meaning as researchers ask questions and pose problems, and children talk about their mathematics and explain their actions" (Hunting, 1997, p.146). In this way, researchers have an opportunity to gain greater insights into the cognitive processes of children's mathematical understanding.

In this study, the interviews were intended to address four core areas:

1. Counting, cardinality and conservation of number

First, two cubes were put on the table. Each child was asked how many cubes there were. The number of cubes was then increased several times, the child being asked each time how many cubes there were. If a child did not count the cubes, the researcher asked the child how they knew the number of cubes in front of them - did they 'just know' (subitise) or use some other method?

The children's understanding of conservation of number was investigated by starting with a number of cubes on the table. Each child was asked how many cubes there were. The cubes were then rearranged and the same question asked.

2. Questions concerning "how many more?"

The tasks the children were asked to undertake began with them being presented with two pots, each containing a different number of cubes (say two cubes in one pot and four in the other). The child was then asked which pot had more. If they had no difficulty with this question, they were then asked how many more cubes were in that pot than in the other. This was repeated a few times.

3. Addition and subtraction

Using a strategy based on Hughes' (1986) Box Task, the interviewer placed a small number of cubes in a box. The child was asked how many cubes were in the box. The box was covered up and one or two cubes were added. The child was asked again how many cubes were in the box. This was repeated a few times before the child was asked to represent the calculation on paper. This procedure was repeated with cubes being taken *out* of the box.

4. Division in the context of sharing

This item was included as an extension of the Box Task and served two purposes.

The first was to see how children shared cubes out (either one at time or in groups). The second was to investigate how they dealt with sharing when fractions were involved (e.g. when sharing 7 cubes between 2 people).

Each interview lasted for up to 25 minutes, depending on the child's responses and level of concentration. Where a child showed confidence and interest in a task, that area was explored in greater depth. The interviews were audio-taped and transcribed for analysis.

DISCUSSION OF INTERVIEWS

Analysis of the interviews presents a picture very different from that suggested by the raw test scores alone. The interviews rest on spoken language: Terezinha Nunes has observed:

> Language is clearly connected to many action schemas: for example, when you ask children "How many ...?", they count. (Nunes, 1996, p. 74)

The child who scored least on the pre-test, for example, (a member of the MLD group) whom we have called 'M1', seems to be demonstrating this tendency as the interview begins:

1.	Interviewer:	How much is that? [puts two cubes on the table]
2.	M1:	One ... two.
3.	Interviewer:	And how did you know that? How did you know it was two?
4.	M1:	Get two and two make two. Get cubes and count it altogether.
5.	Interviewer:	How much is that? [puts four cubes on the table].
6.	M1	One ... One, two, three, four [points to the cubes as he counts].

We cannot be confident that M1 is aware that the last number he utters is a measure of the size of the set he is 'counting'. The following fragment, however, seems to indicate appreciation of both cardinality and conservation of number:

12.	M1:	You give me ... You give me ... one, two, three, four, five, six. [points to the cubes as he counts].
13.	Interviewer:	Right ... and ... if I take them and throw them down again, [picks up cubes and puts them down on the table again] how many are there?
14.	M1:	Six.
15.	Interviewer:	How do you know it's six?
16.	M1:	I said count it altogether ... 'cos its all six.
17.	Interviewer:	OK, and if I do it again, [repeats action] how many have you got?
18.	M1:	Six.

The Year 1 child with whom M1 was paired (here named 'P1') outscored him on the *Basic Number Diagnostic Test*, however, as the following extract shows, his understanding of how numbers work, specifically conservation of number, is less well developed.

9.	Interviewer:	Good [responding to previous answer]. And how much is that? [puts eight cubes on the table]
10.	P1:	One, two, three, four, five, six, seven, eight. [points to the cubes as he counts].
11.	Interviewer:	Right, if I take these eight and put them on the table again, [picks up the eight cubes and puts them back on the table again] how much will you have this time?
12.	P1:	One, two, three, four, five, six, seven, *eight* [points to the cubes as he counts]
13.	Interviewer:	And if I do it again, how much do you *think* we'll have this time when I put them down on the table?
14.	P1:	Don't know.

With an appreciation of cardinality and conservation of number it could be argued that M1 has achieved a greater level of mathematical sophistication than P1. It might be reasonable, therefore to hope that this difference might be reflected in their scores for the *Basic Number Diagnostic Test*. However the impression given by these interview extracts conflicts with the impression given by the Test results - where P1 scored 13 compared to M1's 9. Further study of M1's interview illustrates an area of difficulty which the raw test score alone cannot..

19.	Interviewer:	Right ... How much is that? [puts 15 cubes on the table].
20.	M1:	One, two, three, four, five, six, seven, eight, eight, nine, ten, eleven, twelve, thirty-one, thirty-three. [points to cubes as he counts.]
21.	Interviewer:	How much?
22.	M1:	Thirty-three.

M1 scored poorly because beyond 12 his ability to recite the 'standard number word sequence' (Steffe *et al.*, 1983, p. 25) broke down. What is fascinating here, however, is (a) the fact that he utters precisely 15 numerons' or counting 'tags' (Gelman and Gallistel, 1978, p. 80), and (b) that, in his 'count', the successor of twelve is thirty-one! Is it fanciful to imagine that he 'sees' the numeral for twelve and then makes the common error of reversing the digits of thirteen? Although his understanding of how numbers behave is better developed than P1's, his inability to count conventionally beyond 12 will need to be addressed.

Another MLD child, M3, scored well on addition questions but was unable to add three and one by counting on.

76.	Interviewer:	OK. Now, if we've got those three in there [points to the pot with the lid on] and I put one more in [puts one cube into the pot] how much is in there? If we started off with three and put one more in, how much is in there?
77.	M3:	Four.
78.	Interviewer:	How do you know that?
79.	M3:	I know it is four ... with my fingers.

M3 demonstrated his method by making groups of one finger and three fingers on different hands and then 'counting all'.

Compare with child P3, with whom M3 was matched:

41.	Interviewer:	Right ... This time ... How much is that? [lifts the pot to show the five cubes]

42.	P3:	Five.
43.	Interviewer:	[puts the pot back over the cubes] This time I'm going to put two cubes in. [puts two more cubes under the pot] How much is in there now?
44.	P3:	Seven.
45.	Interviewer:	How did you work that out?
46.	P3:	Thinked.
47.	Interviewer:	I know, but what did you think? Can you tell me how you did it?
48.	P3:	... I know numbers.

P3, who scored less than M3 on the *Basic Number Diagnostic Test* appears to have a better grasp of number relationships than M3.

When it comes to subtraction the differences are even more marked:

67.	Interviewer:	How many were in there to start off with?
68.	P3:	Five.
69.	Interviewer:	No. Let's start again. [removes pot and puts six cubes on the table] How much is that?
70.	P3:	Six.
71.	Interviewer:	Right. Now, I'm going to cover them up [covers the cubes with a pot] and I'm going to take *one away* [takes one cube out from under the pot]. How much is that?
72.	P3:	That's a take away. Take away one.
73.	Interviewer:	And how much is that?
74.	P3:	Five.
75.	Interviewer:	Right. How much have you got in there? [lifts the pot]
76.	P3:	Five.

Compare P3's understanding and application of subtraction to M3's performance:

88.	Interviewer:	Now, [*name*], we're going to do a similar thing again. [puts four cubes into a pot]. How many cubes are in there?
89.	M3:	One, two, three, four [points to the cubes as he counts]
90.	Interviewer:	[Puts lid on] This time I'm going to take one *away*. How much is left?
91.	M3:	Zero.
92.	Interviewer:	So you think there's nothing left in the pot?
93.	M3:.	No [an assenting 'No'].
94.	Interviewer:	Shall we have a look?
95.	M3:	Yeah [takes the lid off]. *Three.*

This child was observed during the testing phase during which he insisted on the use of bricks and a 'count all' strategy to solve the sums 1 + 1; 2 + 1; and 3 -1.

The reason for this mismatch between the impressions given by test scores and interviews is, we suggest, that neither the *Basic Number Diagnostic Test* or the *Basic Number Screening Test* are sensitive enough to analyse mathematical *processing* skills.

FINAL COMMENTS

The inevitable conclusion from this exploratory study is that standardised tests, designed specifically for use with children who are deemed to be slow to develop 'age-appropriate' mathematical competency (Gillham, 1996) are of limited value in terms of the raw test scores alone. This brings into question the practice of applying such tests where raw scores are deemed to be indicative of mathematical competence expected within a particular age-range.

Analysis of the clinical interviews and the strategies used by the children in the tests have made it clear to us that, in most cases, the 'matched' children were, in fact, *mismatched.*

The findings of this study raise questions about the usefulness of the tests used as part of the regular monitoring of students' mathematics progress at the MLD school. Daniels (1988, p. 11) suggests that "there has been little progress towards developing devices which yield information as to *how* understanding has been achieved [in mathematics]". The clinical interview is a 'device' which seems to yield more useful diagnostic information on the character and structure of such understanding.

BIBLIOGRAPHY

Daniels, H. (1988). Misunderstandings, Miscues and Maths, *British Journal of Special Education,* 15 (1), pp. 11- 13.

Daniels, H. (1993). Coping with mathematics in the National Curriculum: Pupil strategies and teacher responses, *Support for Learning,* Vol 8 (2), pp. 65 - 69.

Gillham, W. (1996). *Basic Number Diagnostic Test. 2nd Edition.* London: Hodder and Stoughton.

Gillham, W. and Hesse, K. (1976). *Basic Number Screening Test.* London: Hodder and Stoughton.

Ginsburg, H. P., Kossan, N. E., Schwartz, R. and Swanson, D. (1983) 'Protocol methods in research on mathematical thinking', in Ginsburg, H. B. (ed.), *The Development of Mathematical Thinking*, New York Academic Press, pp. 7-47.

Haylock, D. (1991). *Teaching Mathematics to Low Attainers, 8 - 12.* London: Paul Chapman Publishing.

Hunting, R. (1997). Clinical Interview Methods in Mathematics Education Research and Practice, *Journal of Mathematical Behavior,* 16 (2), pp. 145 - 165.

Hughes, M. (1986). *Children and Number* Oxford: Blackwell.

Nunes, T. (1996) 'Language and the socialisation of thinking' in Puig, L. and Gutierrez, A., *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education, Volume 1* University of Valencia, Spain, pp.71-76.

Steffe L. P., von Glasersfeld, E., Richards, J. and Cobb P. (1983) *Children's Counting Types: Philosophy, Theory and Application*, New York, Praeger.