Key Stage 3 Statutory Assessment 1998 - Report on a *ATL/ATM/NATE* Joint Study

Sue Pope, Roehampton Institute London

The Key Stage 3 tests in mathematics are now well established, and QCA's annual evaluations suggest that they are generally well regarded. In this study, initiated by the curriculum committee of the Association of Teachers and Lecturers, the views of teachers suggest that whilst the tests for mathematics are broadly accepted, there are still a number of issues that need to be addressed: compatibility with GCSE, transparency about which level of attainment each question targets, mental mathematics (time for questions, differentiated papers), how returned scripts are used and the status of teacher assessment.

Introduction

Following the Association of Teachers and Lecturers' (A TL) studies of statutory assessment in 1995 and 1996, and the study undertaken on ATL's behalf on the validity of the key stage 2 (KS2) tests in 1996, ATL initiated a joint study with the Association of Teachers of Mathematics (ATM) and the National Association for the Teaching of English (NATE) on the KS3 statutory assessment 1998. This study was designed to investigate teachers' views of the 1998 statutory tests in English and mathematics at key stage 3.

In addition QCA commission an evaluation of the statutory assessment annually. In 1997 116 schools responded to a questionnaire about the mathematics tests at KS3; they indicated general satisfaction with the tests, however the extension paper was seen as very hard, and little use was being made of returned papers. In 1998 just 63 schools responded to the evaluation survey. Again the level of satisfaction with the tests was high, although the majority of respondents felt their teaching had changed as a result of the tests, and there was considerable dissatisfaction about the delay in getting results to schools.

The study was conducted in two parts. The first part was conducted through a questionnaire which was circulated in ATL, ATM and NATE mailings in June 1998. The questionnaire was divided into eight sections, and included subject-specific questions on the English and mathematics tests. The second part was conducted through two focus groups, one for English, the other for mathematics, involving teachers who had taught the relevant National Curriculum programmes of study and had administered the key stage 3 tests.

A total of 543 questionnaire responses were received by the closing date of 24 July: 229 for mathematics; 314 for English. The coding and inputting of the questionnaires was undertaken by Opinion Research Corporation International, and the data was analysed by ATL's policy unit using an SPSS package. A significant number of teachers responded positively to the invitation to include additional comments on the questionnaire (148 English teachers and 110 mathematics teachers). This paper reports on the mathematics findings only.

Findings

The findings are based on an analysis of:

- the key stage 3 test papers and mark schemes for 1998
- the 229 questionnaire responses, of which 110 had additional written comments
- the issues raised when a focus group of teachers (4) with a sample of their marked scripts (at least 12) met with the researcher for a day.

30% of the respondents were members of ATM. There was no statistically significant difference on any of the questions asked in the responses from A TM members compared with the members of A TL.

Quantitative Responses

Teaching and learning

- 27% of respondents agreed that the tests reflect their view of good mathematics teaching, 35% disagreed.
- 59% of respondents feel under pressure to teach towards the test, 26% do not.

- 37% of respondents said that their teaching style had not been affected by the style of the key stage 3 tests, 40% said their teaching style had changed.
- In response to the proposition that 'too much of my lesson time is spent preparing for the tests' 37% agreed and 42% disagreed.

Students' knowledge and understanding

- More than half of the respondents believe that the tests have not helped to improve their students' depth of understanding (56%), or breadth of knowledge (59%) in mathematics.
- Less than one respondent in six believes the tests have helped to improve students' depth of understanding (15%) or breadth of knowledge (17%).
- Less than half (47%) the respondents believe that the tests provide an accurate reflection of a student's overall ability, whilst 35% believe the tests do provide an accurate reflection of ability.
- 51 % of respondents agreed that the tests adequately represent the national curriculum mathematics programme of study, whilst one in four disagreed. This response is a cause for concern, as much of the national curriculum programme of study cannot be assessed using timed written tests, in particular using and applying mathematics, the use of leT, exploring data in a meaningful way, etc. It appears that what is tested dictates what is perceived to be the national curriculum and consequently what is taught.

Marking and administration

- 42% of respondents agreed that the mark scheme was well designed, 22% did not agree.
- 57% agreed that the external marking was accurate, whilst 18% did not. This is a surprisingly high proportion, as on the focus day several marking errors were found which resulted in a change of level. Some questionnaires were received before marked scripts were returned to schools, so it may be that teachers were not commenting specifically on the 1998 tests, but on their previous experience of the quality of marking.
- 61 % of respondents agreed that the administration of the tests was manageable, and 21 % disagreed.

The mathematics tests

• 57% agreed that the mental arithmetic test was fair, 32% disagreed.

Of the respondents who chose to make additional comments, 25% of those who agreed that the mental test was fair raised the following issues:

there is insufficient time allowed to answer questions
there are only two tiers of test (level 3 to 5, then level 4 upwards) - the tiering should match the tiering of the written papers
the weighting of the mental arithmetic test score is too great when the level is calculated
the administration presents difficulties for pupils with special educational needs.

- 52% agreed that the items on each of the different tests were appropriate, whilst one quarter of respondents. disagreed.

Summary of the Written Comments on the Questionnaires

48% of the respondents chose to make additional comments on the tests. These teachers held similar opinions to the other teachers in that their responses to the mathematics-specific questions on the questionnaire did not differ significantly from the overall responses. Whilst the proportion of written responses from ATM members (53%) was slightly greater than 48%, this is not statistically significant. There was no significant difference in the quantitative responses from ATM members who made written comments, and those of all other respondents.

There were a total of 180 different written comments, with respondents giving between one and four separate points. The comments were classified as follows:

Mental arithmetic test
There were 48 comments about the mental arithmetic tests. The issues raised were:

- having only two tiers meant that many students took a test that was inappropriate. Several respondents requested that the tiering of these tests was the same as for the written tests: *'it is very unfair to ask pupils who are taking the* 4-6 *papers to sit the same mental arithmetic test as those taking the level* 6-8 *paper - there is very little overlap in material'*
- the response times do not give pupils time to think, and pupils with English as an additional language are particularly disadvantaged: *'the test is too time pressured rather than a test of ability'*
- the weighting of the mental arithmetic test is too great
- the two higher tier papers were not of comparable difficulty

• the content of the test was mathematics, and not just arithmetic.
There was relief that jottings are no longer penalised.

Test structure
There were 55 comments about the structure of the tests including:

- over one-third of the comments on test structure described the tests as too linguistically demanding, particularly on the level 3-5 paper
- one-quarter of the comments were about the content of the tests, in particular: Venn diagrams, the context for simultaneous equations, the difficulty of the number operations on the level 3-5 paper 1, the proportion of shape and space on the level 5-7 paper 2, the amount of algebra on paper 1, the perception that the level 5-7 papers were much more demanding than the level 4-6 papers
- one quarter of the comments were about the incompatibility of the key stage 3 tests and GCSE: *'1 cannot believe that level* 7 *is the same as GCSE grade* C.' It is worth noting that the styles of the two assessments are quite different. It was not apparent from the comments whether GCSE was perceived to be more difficult, or otherwise. The principal complaint was that the key stage 3 tests are not sufficiently similar in style to present a useful indicator on the way to GCSE. They represent a diversion from a coherent 11-16 continuum.
- the questions were too long
- the use of shading and different fonts was confusing
- a clear space for the final answer (as on GCSE papers) would be preferable
- the questions make extensive use of paper: *'not typical ... because offinancial restrictions'*
- there needs to be a level 1-4 test
- the tests are well designed and give good coverage of content.

The impact of the tests
There were 27 comments on the impact of the tests:

- one-third concerned the narrowing of teaching and learning in mathematics and teaching to the test: *'lots of practice is needed so students feel adequately prepared',' 'valuable teaching time is lost preparing for the tests'* .
- one-quarter raised concerns about the reliability and validity of the tests: *'As a means of raising standards, they are misguided. The tests will have a long term negative impact on the quality of learning. Formal tests are not the only form of assessment and are certainly not the way to raise the quality of mathematics teaching.'*
  '1 do not think they are necessarily afair reflection of a pupil's capacity ... '
- the timing of the tests: *'having worked towards tests in early May, it is difficult to motivate students for the rest of the academic year'*
- they motivate students: *'particularly boys at the end of Year* 9'
- they inform teaching
- they are used as a basis for testing at the end of years 7 and 8.

Administration

The remaining comments were concerned with the administration of the tests, including marking. Over half of these comments (from one in four of those who made written comments) were about the delay in reporting levels to schools: *'This slippage would not be tolerated from schools.*
Other comments included:

- the difficulty of administering the tests to students with special educational needs: *'Seven members of staff to invigilate for* 89 *pupils in six different venues'*
- the disruption to school routine
- the timing of the tests in the school year
- different mental tests on different days
- assessing pupils whose attainment is below level 3
- the apparent change in threshold marks for the different levels
- concern at the low threshold marks for different levels. To achieve level 6 from the 4-6 papers, just 92 marks out of 150 (i.e. 61 %) were required, and to achieve the target levelS from the same paper 63 marks were required (i.e. 42%). This is of particular concern since the papers are supposed to be structured with a third of the marks for each different level, so on the 4-6 papers there should be 100 marks on levels 4 and 5 (i.e. 67%). The 5-7 paper is similar with 62 marks for the target level, and 99 marks for level 7. Only on the 3-5 and 6-8 tiers are the thresholds for the highest levels in excess of 100 (107 and 109 respectively) whilst thresholds for the target level are still below 50% (72 on both papers).
- the time taken for papers to be marked
- the quality of the marking: *'The marking of Papers* 1 *and* 2 *seemed to have been checked but not the mental arithmetic - one of my pupils had* 16 *marks - not* 10 *as written, because the marker had only added the marks on one side of the paper.'*
- whether or not the testing regime represents good value for money.

Analysis of the Question Papers

The analysis of the tests was undertaken in conjunction with the focus group which comprised four experienced teachers with a sample of scripts. The overall coverage of Attainment Targets 2-4 on the test papers was 2: 1: 1 for level 4 and upwards. There was a small but statistically significant deviation from this weighting in favour of Number and Algebra for the level 3-5 tests.

Mental Arithmetic Tests

Two aspects of the tests were warmly welcomed:

- the mental imagery questions - the title mental mathematics would be more appropriate
- the fact that jottings are no longer penalised.

Concern was expressed about:

- the lack of tiering of the mental tests
- that each mark had equal weight with the marks on the written tests
- the arbitrary 5, 10 and 15 second time limits
- similar questions on the two higher tier papers with different time limits
- the comparability of the two higher tier papers
- the number of questions at the different levels on each paper (level 3,4,5, expected 10: 10: 10 was actually 3:16:10 and 1 level 6; level 4,5,6,7 expected 7:8:8:7 was actually 5:10:10:5 (A) and 4: 13:6:7 (B»

There are two serious implications here:

- for the lower attainers, taking the test willbe a particularly negative experience, as there will be so few questions appropriate to them

- the threshold marks for levels need to be low, e.g. 6 marks for level 3 on test C, when 10 marks should be readily accessible, and just 5 marks on papers A and B for level 4.

Paper 1 (non-calculator)

The following points were regarded as strengths of these papers:

- all calculations are presented horizontally
- the use oflslamic designs in the 'Symmetry' question
- the use of examples and diagrams to facilitate access (e.g. 'Magic Squares', 'Headwork')
- the explanation in the mark scheme about the use of a comma as a decimal point
- questions which encourage students to think (e.g. 'Tiles')
- questions which ask students to explain their reasons (e.g. 'Values').

Concern was expressed about:

• the limited content available for assessment at the higher tiers

- the number of questions which included material beyond the targeted levels
- the use of cultural specific contexts (e.g. 'Jigsaw')
- the amount of reading some questions required.

Of particular concern is the number of questions which are too hard for the paper they are on. Not only is this demoralising for pupils, it could explain why threshold marks for the different levels are so low. The re-introduction of assigning levels to questions is strongly urged. By so doing, teachers would be better informed and, when working through the tests, pupils could decide whether or not it is worth persisting with a question given its level.

Paper 2 (calculator)

A similar analysis was conducted for the calculator paper. Again the goal posts appear to be moving by using some questions which are too hard for the level of the paper. It could be suggested that paper I is being made harder than paper 2 so that the marks from paper I are lower than those for paper 2 in order to support the view that pupils cannot cope without calculators.

General observations

Teachers believe it is important that all students do the tests, to avoid any stigmatising. In some schools, the tests are administered in sections to students with special educational needs. They are concerned that there are gender differences in the test entry: girls who are likely to achieve level 4 take the level 3-5 papers, whilst boys likely to achieve level 4 are entered for the 4-6 papers. Being entered for the level 4-6 papers can boost some students' self esteem, but for others tackling a paper where much of the content is beyond their experience could seriously undermine their self confidence. Teachers are managing the entry to different tiers with sensitivity.

Teachers can feel under pressure to enter students for the extension paper, as managers in school often do not realise that the extension paper for mathematics goes beyond level 8, and is not comparable with the arrangements for English and science. The extension paper cannot be cost effective when teachers prefer to enter students early for GCSE rather than use it. It was noted that, in Wales, there is no extension paper and attainment beyond level 7 is based on teacher assessment only.

When scripts are returned to school there is little evidence of marking other than the numerical marks. It would be helpful if correct work could be ticked so that when students see their papers they can easily see where they have been successful. Some teachers do not let students see their papers, as they feel the gap between taking the tests and receiving the marked scripts is too long for their potential for teaching to be realised.

The current system does not value Teacher Assessment (T A) sufficiently. The final level should be an aggregate of the teacher assessed level for Using and Applying Mathematics and the result of the tests. There is genuine concern that teachers are under pressure to ensure that their T A matches the test

result. If there is a mismatch, they may be perceived as under- or over-estimating a student's performance, rather than recognising that the T A is broader and takes into account things which timed, written tests cannot assess.

Conclusions

As noted in the analysis of questions on papers 1 and 2 there has been a gradual moving of the goal posts so that material which is inappropriately hard is now included in the tests. The non-calculator paper is harder than the calculator paper. This could be remedied by reinstating the referencing system used in the first statutory tests which linked each question to the attainment target and level it was designed to assess. This would also result in higher threshold marks for the different levels, and students having more opportunity to show what they can do.

The mental arithmetic tests should be renamed as mental mathematics tests. There should be fewer questions, with a longer time for each question, and the tiering arrangements should be the same as those for the written papers.

The marking by external examiners should be more rigorously checked. Marks should be made on the papers to indicate where work is right or wrong. The educational potential of the tests would be enhanced if scripts could be returned to school more quickly, and teachers were given the opportunity to inform their practice by systematically reviewing the performance of their students on the tests.

The tests need to be examined more thoroughly to ensure they reflect standard conventions and best practice in mathematics teaching. For many teachers, the tests serve as exemplars which influence their teaching.

It is recognised that finding the right balance between questions in context, the amount of reading required, and context-free questions is not easy. Nevertheless, questions which challenge students to think and to explain their reasoning are welcomed and should be retained.

The issue of compatibility with GCSE requires further exploration. There are some schpols that enter students in year 9 for GCSE. It would be interesting to know if the outcomes are comparable.

The status of Teacher Assessment requires reconsideration. Given the high status of tests for target setting etc. it would be unreasonable to expect teachers to submit T A without seeing the statutory test results beforehand.

References

ATL (1996) *Doing Our Level Best - An Evaluation of Statutory Assessment in* 1995 London: ATL A TL (1997) *Level Best Revisited* London: A TL

ATL (1998) *An Evaluation of the* 1998 *Key Stage* 3 *Tests in English and Mathematics* London: ATL Avon Curriculum & Assessment Support Service (1999) *The Evaluation of Key Stage* 3 *Statutory Assessment Arrangements* 1998 London: QCA

Close, G., Furlong, T. & Simon, S. (1997) *The validity of the* 1996 *Key Stage* 2 *Tests in English, Mathematics and Science* London: A TL

DfE (1995) *Mathematics in the National Curriculum* London: HMSO

SCAA (1995) *Exemplification of Standards in Mathematics: Key Stage* 3, *Levels* 4 *to* 8 London: SCAA

Social & Market Strategic Research (1998) *The Evaluation of Key Stage* 3 *Statutory Assessment Arrangements* 1998 London: QCA

ACKNOWLEDGEMENTS

We owe a debt of gratitude to the 543 teachers who took the time, whilst under considerable pressure themselves, to complete the questionnaire, many making additional comments. Special thanks are due to the teachers who participated in the focus groups, and to their headteachers for releasing them from school for the day. Finally, our thanks go to colleagues at the Qualifications and Curriculum Authority for their willingness to share issues of professional concern. While we may not always agree, the dialogue must continue.

Anne Barnes (NATE), Sheila Dainton (ATL), Gwen Evans(ATL), John Moss(NATE), Sue Pope(ATM)