

## Making Good Progress MEI response to DfES consultation

### Introduction

“Making Good Progress” recommends one test per National Curriculum Level and using the proportion of students who progress at least two levels over a Key Stage as a performance measure.

This response has four sections:

- A review of the assumptions underlying these recommendations
- An examination of the technical issues arising from measuring Test Levels
- A description of the issues which progression measures would raise
- A commentary on the pedagogical issues that would be raised.

### Assumptions

The proposals presented in “Making Good Progress” depend on three underlying assumptions and require all three to be valid. However, it is far from clear that this is the case for any of them.

#### ***There has been a considerable improvement in standards***

“Making Good Progress” begins with the statement: “*There have been enormous improvements in recent years in the standards which pupils achieve at school.*”<sup>1</sup> While we would like to believe this to be the case, we are aware of evidence that casts doubt upon it and, particularly, the use of the word “enormous”. The IPPR report, “Assessment and Testing”<sup>2</sup>, presents evidence to suggest that improvements in National Curriculum Levels overstate underlying improvements in attainment.

*“Although the two are not directly comparable, improvements in TIMSS (Trends in International Mathematics and Science Study) are thus much less impressive than the measured improvements in key stage test results. The Statistics Commission considered these issues in 2005 and concluded that: ‘The Commission believes that it has been established that (a) the improvement in Key Stage 2 test scores between 1995 and 2000 substantially overstates the improvement in standards in English primary schools over that period, but (b) there was nevertheless some rise in standards.’ (Statistics Commission 2005: 4). Looking at the secondary phase, the percentages of pupils attaining the benchmark at Key Stage 3 and Key Stage 4 have continued to rise although progress on international attainment measures has stalled. Evidence from TIMSS for Key Stage 3 (Year 9) does not show any significant change in performance between 1995 and 2003 (Ruddock et al 2004). Analysis of the international study PISA (Programme for International Student Assessment) shows that for a given score at Key Stage 3 or Key Stage 4, pupils attained on average a higher PISA score in 2000 than in 2003 (Micklewright and Schnepf 2006).”<sup>3</sup>*

### *Causality*

The questionnaire accompanying “Making Good Progress” asks:

*“The document asks whether - without compromising the framework of tests, targets and performance tables which have helped drive up standards so sharply over the past decade - we can adapt the system to support a focus on **progress as well as absolute standards.**” Do you agree?”<sup>4</sup>*

This makes a strong causal connection between testing (and associated targets and performance tables) and improvement in standards; this is no more than an unproved assumption. Other initiatives to improve teaching and learning, such as the National Strategies, could be mainly responsible for any improvement in standards.

The relationship between monitoring and standards is complicated and the effects are not always wholly beneficial. The introduction of high-stakes assessment has resulted in a narrowing of the curriculum experienced by many students as their teachers feel the need to teach more closely to the test to ensure that their pupils achieve targets, as measured by the end of Key Stage Tests.

### *Level definition*

The assumption that National Curriculum Levels are well-defined is crucial to the possibility of writing single-level tests. However, the 2005 analysis of two-tier GCSE found that the target level of questions did not correlate well to the final grades of students who were successful in answering them. *“We found on the Pilot that our question level data (Section 3) challenged the assumption that particular questions could be precisely targeted at particular grades.”*<sup>5</sup> This means that it may not be possible to construct mathematics questions which discriminate reliably between students at a particular level (and above) and those at a lower level.

The intention of the original design of the Levels in the late 1980s, involving fourteen attainment targets, was to provide a picture of a student’s aptitude and progress over all aspects of mathematics. This proved unwieldy and was transformed into the more simplistic concept of a single average level. The limitations of such an average need to be recognised and, in particular, the serious errors that will arise if differences in Levels are used as a basis for measuring progress across a Key Stage.

## Technical issues arising from measuring students' levels

There are two ways in which a student can currently be assigned a level:

- teacher assessment (Teacher Level)
- multi-level tests (Test Level)

End of Key Stage Test and Teacher Levels can be different. If a student being "at Level 7" in mathematics is regarded as a well-defined concept, similar in nature to being a certain height, all measurements should lead to the same result. However, mastery of mathematics has more in common with the ability to play a musical instrument or sporting skill; different measurements of such abilities can yield differing results. Appendix 1 describes how Teacher Levels and Test Levels are currently ascertained.

The fact that Test and Teacher Levels can be genuinely different is unhelpful, inviting the public to question teachers' judgement. This situation would be made worse by the introduction of single-level tests, unless they were a replacement for multi-level tests. Single-level tests would involve a third method of assessing a student's level and so, inevitably, could give a different result to the other two.

In a single-level test at Level 7, most, and possibly all, of the material will be at Level 7. This contrasts with the KS3 Level 5-7 test, where only  $\frac{1}{3}$  of the test is at Level 7. It will no longer be as easy to lead students into questions appropriately and so the Level 7 test will be harder. In the current Level 5-7 test, it is usually possible for students to be awarded a Level 7 by scoring nearly full marks on the Level 5 and 6 questions.

There are two consequences that need to be understood:

- The resulting national distribution of Test Levels based on single-level tests would differ from the current one, suggesting a decline in standards of attainment.
- It will not be acceptable to have a situation in which it is possible to choose the test that will give the most favourable result so the two types of test cannot co-exist.

## Issues arising from Progress Targets and the Progression Premium

*"The reason for pursuing higher standards is not in order to achieve numerical targets or to deliver accountability. Useful and necessary as these are, they are the servants and not the masters. The data and targets we set are the means towards the objective of equipping pupils with the skills and knowledge they need: education for self-fulfilment, access and equality. So it is important that we use our data and set our accountability targets to achieve the ends we most value."*

(DfES, 2006)<sup>6</sup>

The use of targets for schools based on increasing the number of students making two levels of progress over a Key Stage would only be fair if the following assumptions were reasonably accurate.

- Levels can be measured sufficiently accurately for the difference in levels to be a meaningful measure.
- It is possible for all students to make two levels of progress in a Key Stage.

***Can levels be measured sufficiently accurately to be used as a basis for measuring progress?***

Recent work by Dylan Wiliam has cast doubt on the accuracy of results at KS2 and KS3, estimating that about 32% of KS2 results and 43% of KS3 results are at least one level out.<sup>7</sup>

In “Making Good Progress” it is proposed that a secondary school’s performance should be judged on the basis of the difference in its students’ levels at KS2 and KS3. However, the errors in these differences will be greater than the errors at either Key Stage and so will be very substantial. This cannot provide a fair basis for paying money to schools.

Recognising the progress of lower attainers is an excellent aim but the evidence is against progression targets based on Test Levels being an appropriate way to do so.

***Is it possible for all students to make two levels of progress in a Key Stage?***

The National Curriculum Levels were designed with the intention that an average student should make one level of progress every two years; this is the historical basis for the “the national expectation” of a gain of one level for every two years. However, it was never expected that all students would proceed at the same rate. In practice, some progress one level per year, others take four or five years to move to the next Level. Variability in students’ rates of learning is a fact of life. It is to be expected that students who have found a subject difficult will tend to progress more slowly than average; this is borne out by the statistics.<sup>8</sup>

A further complication is that each National Curriculum Level covers a range of attainment; a student who starts a Key Stage at the top of the Level 5 spectrum will find it easier to reach Level 7 over the Key Stage than a student at the lower end of Level 5. Many schools use National Curriculum sublevels to track student progress; however, the use of such sublevels in national progression measures would be problematic due to the difficulty of assigning such sublevels accurately.

The proposed progression premium will be payable to schools who increase the proportion of low-attaining students gaining two levels during a key stage. It is manifestly unfair on schools, teachers and, above all, students to introduce a system of payment dependent on outcomes that cannot be measured reasonably accurately nor can be reasonably achieved by many students.

At KS3, “*the expected level of progress is somewhere around one and a half levels.*”<sup>9</sup> A student who makes one test level of progress could be making “expected progress”; two test levels is beyond this. To make low-attaining students, who already know they are below average, feel an even greater sense of failure because they have not improved by two levels will do nothing to achieve “*education for self-fulfilment, access and equality*”.<sup>10</sup> It would be tactically and morally wrong if targets resulted in failure to recognise students’ progress; to do so just before they move onto GCSE is to encourage a sense of inadequacy which will demotivate students as they begin to work towards nationally recognised qualifications.

## Impact on pedagogy

Our concern is not with single-level tests in themselves but with the way in which they are used. There have been examples of single-level tests that have worked reasonably well, for example, GAIM and the MEI National Curriculum scheme for GCSE.<sup>11</sup>

In the MEI scheme, the level tests formed part of the GCSE. A few teachers also used them at an earlier stage, focusing their teaching on fostering mathematical understanding but using the tests, en passant, to provide students with information about how they were getting on. Performance tables had not been invented in those days and these tests were not being used to judge schools.

There is evidence to suggest that some of the recent improvements in National Curriculum Test Levels are due to teaching to the test; using single-level tests to hold schools accountable would encourage this. A 2003 survey, conducted on behalf of the NSPCC, showed that the biggest cause of worry for 11-16 year olds was examinations.<sup>12</sup> Increasing the frequency of high-stakes testing for schools will increase the pressure on students to perform. In our view, it is unethical to put children under such pressure when the results are of more importance to the school than they are to individual students. Alternative ways should be found to monitor schools that do not use children as pawns in an adult game.

The emphasis on student progress, formative assessment and a willingness to consider different forms of National Curriculum testing in “Making Good Progress” are all welcome. It is true that tests provide national benchmark information about attainment and progress, can provide teachers with information which will help them to teach their students more appropriately and can also motivate students by helping them to see the progress they are making. However, using single-level National Curriculum Tests as a basis for rewarding schools whose students meet progress targets would have implications for teaching and learning which would be highly undesirable; this would seem positively to promote teaching to the test. Students, as well as teachers, would be encouraged to view success in the test, rather than understanding the mathematics, as the aim.

*“To some extent what is needed is a jump from one equilibrium, in which teachers are too busy training their pupils to pass national tests to focus on assessment for learning, to a better one in which teachers make more use of assessment for learning and as a result their pupils perform better in national tests.”*

*(IPPR, Dec 2006)<sup>13</sup>*

### ***Formative assessment and summative testing***

Black and Wiliam surveyed 250 studies which indicated that effective use of formative assessment can approximately double pupils’ rate of progress.<sup>14</sup>

There is no reason to believe that, without a change in the measures used for schools’ accountability, introducing single-level tests would encourage formative assessment; the evidence suggests the opposite; the Primary Assessment, Curriculum and Experience project (PACE) found that, following the introduction of Key Stage Tests, primary teachers’ assessments became more summative.<sup>15</sup>

A move to single-level tests would strongly encourage a move to single-level teaching in order to prepare students for them. This has not been common since the early days of the National Curriculum due to the fragmentation of learning it can lead to. Such a move would be counter to the proposed changes to KS3 and 4 Programmes of Study which allow for an increase, rather than a reduction, in curricular freedom and personalisation; it would undermine existing good practice.

### *Alternative methods of assessment*

The principle of piloting alternative forms of assessment is to be commended; there are several alternatives which might be explored. For example, online testing which adapts to what students are able to do so that they are tested at an appropriate level. This would also allow rapid feedback to schools to inform future teaching.

A national graded system of mathematics tests might be motivating for students. This would need an integrated package of teaching and learning materials in addition to assessment materials. The National Curriculum Level descriptors were written to enable teachers to make judgements about their students' levels of understanding rather than to provide a syllabus for teaching. A considerable investment of time and resources would be necessary to design an appropriate curriculum, with associated support materials, to ensure that such a system was successful in improving learning in mathematics. This is particularly important in view of the number of non-specialist staff teaching mathematics at KS2 and KS3.

The IPPR report "Assessment and Testing"<sup>16</sup> proposes using teacher assessment. This would be moderated through samples of students' work; this sampling would encourage good curriculum coverage. Such a proposal would encourage formative assessment and discourage "teaching to the test" and is worthy of further exploration, including a pilot.

## Summary

- Single-level tests will produce different results from multi-level tests and so these forms of testing cannot co-exist.
- It is neither helpful nor fair to pay a premium to schools based on a measure that is not reasonably accurate and requires performance that cannot reasonably be achieved by many students.

While the present system of testing students to monitor schools remains in place:

- Using single-level tests as a replacement for end of Key Stage Tests would increase the pressure on students and lead to a less satisfactory learning experience.
- A move to single-level tests would lead to a national drop in Test Levels without any reduction in overall attainment.
- The difficulty of assigning Levels to students reliably means that a difference in Levels as a measure of progress for individual students is not fit for purpose. Giving money to schools on the basis of such a measure would be not only unreliable but also reinforce a “teaching to the test” culture.

If, however, a different method of monitoring schools was introduced:

- A system of graded tests in mathematics could be introduced, as part of a coherent curriculum. For such a system to be successful, teachers need to teach in a way which encourages the progression and development of mathematical thinking. Appropriate resources for teaching and learning need to be developed, in addition to assessment materials.

## Appendix 1: Teacher Levels and Test Levels

### *Teacher Levels*

At the end of a Key Stage, teachers assign a Level to each student in each of the four attainment targets and use a weighted average to produce an overall Level. It is common for a student to be performing at differing Levels in different attainment targets, so students who achieve the same Teacher Level may have quite different strengths. Moreover, as Teacher Levels are rounded to the nearest whole number, students reported at Teacher Level 6, for example, could have unrounded Levels from 5.5 up to 6.5.

### *Test Levels for the current system*

Students sit two written papers in mathematics, one without a calculator, the other with a calculator; they also take a mental mathematics test in which they write answers to spoken questions. The marks on the written papers of the Level 5-7 tier KS3 Test are typically equally distributed between questions targeted at each of the three levels, (there is a similar pattern for the other tiers). Candidates are awarded Levels on the basis of their total scores. The thresholds for tier 5-7 usually allow a candidate to be awarded a Level 7 by scoring nearly full marks on the Level 5 and 6 questions.

## References

1. *Making Good Progress*, DfES, 2006
2. *Assessment and Testing: Making space for teaching and learning*, IPPR, December 2006
3. *Assessment and Testing: Making space for teaching and learning*, IPPR, December 2006
4. *Making Good Progress (Questionnaire)*, DfES, 2006
5. *Moving to two-tier GCSE mathematics examinations An independent evaluation of the 2005 GCSE Pilot and Trial*, Gordon Stobart, Tamara Bibby and Harvey Goldstein, University of London Institute of Education with the assistance of Ian Schagen (NfER), and Mike Treadaway (Fischer Family Trust), September 2005
6. *Making Good Progress*, DfES, 2006
7. 'The Reliability of Assessments' in *Assessment and Learning*, P Black and D Wiliam, Gardner J (ed), London: Sage, 2006
8. *Making Good Progress*, DfES, 2006
9. *Making Good Progress*, DfES, 2006
10. *Making Good Progress*, DfES, 2006
11. *Delivering Curriculum Pathways in Mathematics*, MEI, 2005
12. *Someone to turn to?*, NSPCC, 2004
13. *Assessment and Testing: Making space for teaching and learning*, IPPR, December 2006
14. 'Assessment and classroom learning', P Black and D Wiliam, *Assessment in Education* 5 (1) 7-74, 1998
15. *What pupils say: changing policy and practice in primary education – findings from the PACE project* A Pollard and P Triggs with P Broadfoot, E McNess and M Osborne, London and New York: Continuum, 2000
16. *Assessment and Testing: Making space for teaching and learning*, IPPR, December 2006