

**HOUSE OF COMMONS
SELECT COMMITTEE ON EDUCATION AND SKILLS**

TESTING AND ASSESSMENT

**SUMMARY OF EVIDENCE
from
THE ASSOCIATION FOR SCIENCE EDUCATION
June 2007**

1. The Association for Science Education welcomes the opportunity to make this submission on *Testing and Assessment* and has consulted widely with its members and others who are drawn from all phases and areas of science education in order to bring together a range of evidence from a variety of perspectives.
2. Although the issues relating to testing and assessment are generally generic and apply across all subjects the impact of some practices has had a particularly detrimental effect on the teaching of science and has, in turn, contributed to the disaffection with the subject expressed by pupils across all phases of education.
3. ASE fully recognises that testing and assessment must meet a range of demands but would argue very strongly that in the current climate the balance is wrong with too much emphasis being placed on the demands for accountability to the detriment of the quality of pupils' learning as a result of a narrowing of the curriculum experienced by pupils. Broadly teachers support the need to redress the balance in order to put more emphasis on formative assessment, often referred to as assessment for learning (AfL).
4. In order to encourage learning pupils' progress needs to be assessed, both in order to help learning (formative) and to report on learning (summative). Using assessment to help learning should be central to education and there is substantial evidence as to the effectiveness of formative assessment.
5. Whilst it is relatively easy to state the essence of the problem, finding solutions is not so straight forward. ASE argues that:
 - a. there is a need to reduce the overall burden of testing and assessment on teachers and pupils as well as to redress the balance between summative and formative assessments;
 - b. steps should be taken to remove the culture of 'teaching to the test' in favour of genuine support for learning through formative assessment approaches;
 - c. greater investment is needed for developing assessment strategies and pedagogy which use a wider range of styles and improved feedback which instils a greater sense of achievement and progress for students;
 - d. teachers and other staff need to be supported, through appropriate CPD, and the necessary time made available in order to develop and implement appropriate processes to ensure the value of testing and assessment is maximised to the benefit of students.
6. With regard to testing and assessment in science specifically, in addition to the above, ASE argues that:
 - a. further consideration should be given to the issue of comparative difficulty of examinations across subjects and actions taken to address the current situation where certain subjects, especially sciences and maths, are perceived to be more difficult;
 - b. the place of 'coursework' and assessment of practical work needs to be revisited in order to reflect the investigative nature of science more strongly.

**HOUSE OF COMMONS
SELECT COMMITTEE ON EDUCATION AND SKILLS**

TESTING AND ASSESSMENT

**A SUBMISSION OF EVIDENCE
from
THE ASSOCIATION FOR SCIENCE EDUCATION
JUNE 2007**

Introduction

1. As stated in its Royal Charter The Association for Science Education¹ (ASE) exists, among other things, to provide 'an authoritative medium through which the opinions of teachers of science may be expressed on educational matters'. ASE therefore welcomes the opportunity to make this submission on *Testing and Assessment* and has consulted widely with its members and others who are drawn from all phases and areas of science education in order to bring together a range of evidence from a variety of perspectives. In particular, in addition to the information from external sources, this submission has drawn on first hand contributions from members of ASE Council, committees and our two Special Interest Groups (National Advisers and Inspectors Group for Science² (NAIGS) and, Association of Tutors in Science Education³ (ATSE)). In addition ASE has contributed to and fully supports the joint statement which will be submitted by the SCORE⁴ partnership.
2. Although the issues relating to testing and assessment are generally generic and apply across all subjects the impact of some practices has had a particularly detrimental effect on the teaching of science and has, in turn, contributed to the disaffection with the subject expressed by pupils across all phases of education. After an overview of assessment and its purposes, the submission which follows presents the key points that have been raised during our consultations in response to the questions set out in the call for evidence.

Assessment: an overview

3. Assessment is a complex issue that is a key element in education but, in its different forms is used to address the, often competing, demands of a wide range of stakeholders. Broadly these demands fall into two groupings:
 - a those which focus on learning and teaching aimed at supporting and encouraging the progress of individual students;
 - b those that are designed to provide 'accountability data' allowing for comparison of the performance of groups and institutions, regionally, nationally and internationally.
4. Although similar assessment approaches might be used the purpose to which they are put will often significantly alter the way in which the assessment is carried out, the way in which the assessment is perceived by students, and the manner in which the outcomes of the assessment are used. For example, a simple test might be used with an individual pupil or a class in order to find out how well they have understood some new vocabulary. The results might then be used by the teacher to plan activities that re-enforce the words

¹ Appendix 1 provides a summary of the aims of The Association for Science Education (ASE)

² Appendix 1 provides a summary of the aims of The National Advisers and Inspectors Group for Science (NAIGS)

³ Appendix 1 provides a summary of the aims of The Association of Tutors in Science Education (ATSE)

⁴ Appendix 1 provides a summary of the aims of the SCORE partnership

in appropriate contexts, thereby helping pupils not only to remember the words but also to gain a better understanding of their meanings. On the other hand a similar test might be used to find out how many pupils have remembered the vocabulary and then the results used to compare one class with another, or the performance of schools in a region or nationally. In the latter case unless something else is done with the information there is no impact on pupils' learning but if used appropriately it might provide evidence of overall standards.

5. This somewhat simplistic example illustrates the key dilemma that exists in the UK, more specifically England, at the present time; i.e. the need to use assessment effectively in order to inform teaching and thereby support pupil learning yet, at the same time, to meet the demands for accountability and public assurance that overall standards are being maintained and, where possible, improved.
6. ASE fully recognises the need to meet both sets of demands but would argue very strongly that in the current climate the balance is wrong with too much emphasis being placed on the demands for accountability to the detriment of the quality of pupils' learning as a result of a narrowing of the curriculum experienced by pupils. There certainly is evidence of this happening in science and, from what teachers have said to us, they feel constrained by the pressure put on them, often in subtle ways, to increase the number of pupils getting Level 5, 6 or 7 in SATS or gaining 5 A*- C at GCSE, as the following quotes illustrate.

"After Christmas we do 3 or 4 mock SATs papers. No one seems to have the guts to just carry on teaching good science."

"Year 10 and 11 just seem to be on a treadmill of one module exam after another. That wouldn't have inspired me to take science (and it certainly doesn't inspire them.)"

7. More positively, teachers have also commented, they would like:
 - *"to be able to continue teaching engaging science in year 6, rather than feel pressured to do endless SATs preparation"*
 - *"assessment to motivate my pupils and reward their success"*
 - *"assessment to encourage learning (and not be a full stop at the end of learning)."*
8. If we are to encourage learning then we must make sure that pupils' progress is assessed, both in order to help learning (formative) and to report on learning (summative). Using assessment to help learning is central to enquiry-based education and there is substantial evidence of its effectiveness.⁵ Similarly, the introduction of thinking skills is known to lead to improved outcomes. In both these cases assessment is part of teaching and there is a lesser role for grading, 'levelling', or making judgements. Whilst it is helpful to learners to have levels, understanding of progression of scientific skills and knowledge concepts is key to ensuring that teachers are able to move children on in their learning. The aim is for teachers and pupils to collect and use information that can be used to decide where children have reached, what their next steps are and how to take them.
9. In the case of summative assessment, we should also be able to report on the whole range of learning outcomes. At present this does not happen, because tests cannot do this. Even though tests (at present) only occur at the end of key stages, there are other, sometimes commercial, tests that are given at other times; and teachers tend to emulate

⁵ Black, P., Harrison, C., Lee, C., Marshall, B, and Wiliam, D. (2003) *Assessment for Learning: Putting it into Practice*. Maidenhead: Open University Press

these tests when making their own. Moreover, as the work of the Assessment Systems for the Future (ASF) reported,⁶ a system based on testing is flawed, because:

- It fails to provide information about the full range of educational outcomes that are needed in a world of rapid social and technological change and therefore does not encourage the development of these skills.
- It inhibits the development of assessment for learning which is shown to raise achievement levels and reduce the gap between higher and lower achieving pupils.⁷
- The data it provides are less reliable than they are generally thought to be. For example it has been estimated that the key stage tests in England result in the wrong levels for at least a third of pupils at the end of KS2 and up to 40 per cent at the end of KS3.⁸
- The weak reliability of tests means that unfair and incorrect decisions will be made about some pupils, affecting their progress both within and between schools⁹ and beyond school.
- There is no firm evidence to support the claims that testing boosts standards of achievement.¹⁰
- It reduces some pupils' motivation for learning.¹¹
- It imposes unnecessary stressful conditions that prevent some children from performing as well as they can.¹²
- It encourages methods of teaching that promote shallow and superficial learning rather than deep conceptual understanding.¹³
- Apart from the anxiety caused by the tests, revision and practise tests take up a considerable amount of time, for example, in years 5 and 6 this is estimated by the ASF project to be the equivalent of three weeks of learning in each year.¹⁴

10. Broadly teachers support the need to redress the balance in order to put more emphasis on formative assessment (often referred to as assessment for learning (AFL) as the following comments illustrate,

"Assessment for learning (AFL) should be the prime business of schools - if you concentrate on AFL, with professional development for teachers to support this, then learning will improve and that will mean that everyone benefits."

"Personal learning planning covers both learning and accountability."

"Good teacher assessment requires teachers who have good pedagogical subject knowledge – there are issues of the need for CPD and teacher confidence in their ability to make judgements. Over reliance on summative tests may lead to de-skilling

⁶ ARG (Assessment Reform Group) (2006) *The Role of Teachers in the Assessment of Learning*. Obtainable from the ARG website: www.assessment-reform-group.org and from the CPA office of the Institute of Education, University of London

⁷ Pollard, A. and Triggs, P. (2000) *Policy, Practice and Pupil Experience*. London: Continuum International Publishing Group

⁸ Wiliam, D. (2001) *Reliability, validity, and all that jazz*. *Education 3-13*, 29(3) 17-21

⁹ Black, P. and Wiliam, D. (2006) *The reliability of assessment*, in J. Gardner (ed) *Assessment and Learning*. London: Sage

¹⁰ Tymms, P. (2004) *Are standards rising in English primary schools?* *British Educational Research Journal*, 30 (4):477-494

¹¹ Harlen, W. and Deakin Crick, R. (2003) *Testing and motivation for learning*, *Assessment in Education*, 10 (2): 169-208

¹² Reay, D. and Wiliam, D. (1999) *'I'll be a nothing': structure, agency and the construction of identity through assessment*, *British Educational Research Journal*, 25: 343-345

¹³ Harlen, W. and James, M. (1997) *Assessment and learning: differences and relationships between formative and summative assessment*, *Assessment in Education*, 4(3) 365-80

¹⁴ Harlen, W. (2007) *Assessment of Learning*. London: Sage

of teachers in assessment techniques, knowledge of level characteristics and ability to plan next steps in learning."

"Summative assessment by testing, with results used for accountability, can lead to less creative teaching and learning, a narrower experience for pupils and pupils with less real understanding of, or interest in, science."

"Skills would be more effectively assessed by using investigative activities with clear assessment criteria and an AfL approach."

"Current assessment is too narrow and does not really assess the important aspects of creativity, thinking etc. at present we are teaching exam techniques".

"Teachers need CPD and continuing support to ensure they have a clear understanding of pupil progress in science skills and concepts."

11. Whilst it is relatively easy to state the essence of the problem, finding solutions is not so straight forward. In part this is because the situation is so complex that there is no single solution nor a simple causal relationship between pupil achievements and the way they are assessed or taught. However, based on information we have gathered some of which is presented in response to the questions addressed in the subsequent parts of this submission, ASE would argue that:

- there is a need to reduce the overall burden of testing and assessment on teachers and pupils as well as to redress the balance between summative and formative assessments;
- steps should be taken to remove the culture of 'teaching to the test' in favour of genuine support for learning through formative assessment approaches;
- greater investment is needed for developing assessment strategies and pedagogy which use a wider range of styles and improved feedback which instils a greater sense of achievement and progress for students;
- teachers and other staff need to be supported, through appropriate CPD, and the necessary time made available in order to develop and implement appropriate processes to ensure the value of testing and assessment is maximised to the benefit of students.

12. With regard to testing and assessment in science specifically, in addition to the above, ASE would argue that:

- further consideration should be given to the issue of comparative difficulty of examinations across subjects and actions taken to address the current situation where certain subjects, especially sciences and maths, are perceived to be more difficult;
- the place of 'coursework' and assessment of practical work needs to be revisited in order to reflect the investigative nature of science more strongly.

Enquiry questions: General Issues

Why do we have a centrally run system of testing and assessment?

Who is the QCA accountable to and is this accountability effective?

What role should exam boards have in testing and assessment?

13. Ultimately the existence of a centrally run system of testing and assessment is the result of a decision by Government that such arrangements should be in place. It would therefore be easy to take a cynical view as to why it was determined that such a system was / is necessary. However, there are advantages to having a centrally run system that endeavours to ensure consistency and comparability of testing and assessment across the country. It also provides the potential to monitor standards and, indirectly, provide evidence on the impact of policy changes that relate to the curriculum and other aspects of educational provision in schools and colleges. Furthermore a centrally run system of testing and assessment should help to provide a clarity and coherence as to the qualifications that are available to students as part of working to meet their aspirations, aptitudes and interests as well as meeting the needs of other stakeholders.
14. The major disadvantage of any system, particularly one that is run centrally, is the potential it provides for distortion of its original purposes so that, for example, monitoring of standards leads to enforced compliance in order to meet targets. This in turn results in a culture that limits innovation and enjoyment of learning (see paragraph 18). Avoiding such distortion of any system requires clear lines of responsibility and processes for monitoring the effectiveness of the procedures against the overall educational aims and objectives. This requires a much closer relationship between the development of the curriculum and the testing and assessment arrangements that are to be in place.
15. If a centrally run system is to work effectively then it is crucial that the body / bodies responsible should have clear transparent lines of accountability and that their independence is not compromised in any way. QCA currently has an overall responsibility for the curriculum, testing and assessment arrangements and reports to the Secretary of State. However the way in which this responsibility is exercised is not entirely clear to the vast majority of stakeholders. Further confusion arises with the establishment of the NAA (a subsidiary of QCA) to undertake matters relating to testing and assessment and also the regulatory role QCA has in relation to the Awarding Bodies. Although ASE enjoys good relations with QCA, which it values, we would argue that **there is a need for greater clarity as to the ways in which QCA carries out its role and that there should be improved mechanisms for engagement with stakeholders in relation to matters of testing and assessment.**
16. **ASE would also argue for a greater clarity in the way in which Awarding Bodies conduct their role in testing and assessment.** In particular there is concern that the changing character of Awarding Bodies, including their publishing activities as providers of educational resources, will lead to conflicts of interest and to compromising of their 'objective status' in relation to the development and implementation of testing and assessment arrangements.

What other systems of assessment are in place both internationally and across the UK?

17. There are trends across the UK towards making greater use of teachers' judgments and less testing for summative assessment. Scotland is well into a programme of reform in assessment, *Assessment is for learning*, developed through collaboration between teachers, researchers and policy-makers. The part of the system dealing with summative assessment of pupils depends on teachers using evidence from regular activities to decide when a pupil has reached a particular level. Quality assurance of teachers' judgments can be through moderation with other teachers or using a test drawn from a bank of assessment tasks. In Wales, where end of key stage tests were finally phased

out in 2005, summative assessment is based on teachers' judgments moderated in inter-school groups at Key Stage 2 and by accreditation of schools at Key Stage 3. In Northern Ireland, too, revisions underway will lead to assessment at the end of Key Stages 1, 2 and 3 taking the form of moderated teachers' judgments. In England, since 2005, teachers have had more responsibility for assessment at the end of Key Stage 1. Thus there are examples of alternative approaches to testing and assessment within the UK that are seeking to redress the balance between formative and summative assessment. ASE would suggest that these are worthy of further consideration, however **ASE is very concerned that recent proposals for England¹⁵ to add single-level tests to be administered twice each year to those ready to take them, will increase rather than decrease the amount of testing.**

Does a focus on national testing and assessment reduce the scope for creativity in the curriculum?

18. In essence the answer is 'Yes'. When a high stakes testing system is in place it is inevitable that teachers will teach what they think is needed to pass the tests. This has several effects. Within the tested subjects it narrows the focus to aspects of the curriculum which are easily testable and reduces the time available for other elements or for creative planning which extends beyond narrow subject boundaries. It creates a culture of compliance, where the child is not at the heart of the process and factual learning has precedence over skills, experiences and interest. Assessment gives messages about what is important e.g. there is no testing for KS1 science which is teacher assessed so end of key stage assessment data is not included in value added calculations. Non tested subjects in primary school are seen as lacking importance resulting in reduced time being given to them, especially towards the end of the key stage. **The full value of a creative, linked curriculum which addresses the interests, needs and talents of all pupils is not exploited because many schools seem to be afraid to innovate when test scores might be affected (even if evidence shows they might actually go up).**

National Key Stage Tests: the current situation

How effective are the current Key Stage tests?

Do they adequately reflect levels of performance of children and schools, and changes in performance over time?

Do they provide assessment for learning (enabling teachers to concentrate on areas of a pupil's performance that needs improvement)?

19. Tests that are designed to be taken by large numbers of pupils, with all those at a certain age or stage taking the same items, are clearly limited in the number and range of items that can be included. The number is limited by the time and length of test that it is reasonable for a pupil to be given. The range is limited because, where large numbers of pupils are tested, items require written answers that can be readily marked as unambiguously as possible. Otherwise they become impossibly expensive to process. These limitations mean that the items included can only be a small sample - and a particularly restricted sample - of the range of items theoretically necessary to cover the domain being assessed to give results of acceptable validity. The reliability of the assessment is also affected since many different selections of items could be made in creating the test, each likely to lead to a different result. A pupil achieving a certain level or score might have a higher or lower score on a different selection of items. Calculations based on Key Stage 2 test results indicate that this effect could result in at least a third of pupils being given the wrong level.¹⁶ This evidence would argue that, National Tests have neither the high validity nor high reliability that they are widely

¹⁵ DfES (2007) *Making Good Progress* Consultation. London: Department for Education and Skills

¹⁶ Wiliam, D. (2001) Reliability, validity and all that jazz, *Education 3-13*, 29 (3) 17-21

assumed to have. They do not give good information about what individual pupils can do and they provide a poor estimate of the level of attainment at the national level because they sample such a small proportion of the relevant goals.

20. Science tests, for example, seem to measure literacy (reading comprehension), maths and logic as much as science which is not necessarily a bad thing. However, in terms of ensuring that pupils are enthused about science, have a genuine depth of knowledge and understanding across the whole subject, can investigate independently and can make connections and apply their knowledge existing tests are limited. Indeed **high scores can mask underlying issues of poor practice in the school**. For example, lengthy and intensive revision programmes can place over-emphasis on knowledge and understanding at the expense of Sc1 (potentially damaging to pupils' motivation and attitudes).
21. Interpretation of the impact and usefulness of tests is further complicated by the confusion between formative assessment that helps learning day to day (which national tests clearly cannot do) and the formative use of summative assessment in medium and long-term planning. Neither current tests nor the proposed single-level tests support formative assessment in the former meaning. Indeed, the pressure they put on teachers, because of their use as targets, prevents attention being given to using such assessment to help learning. Research also shows that frequent testing reduces pupils' motivation for learning¹⁷ and imposes stressful conditions¹⁸ that prevent many from performing as well as they are able – a further source of low reliability of the outcomes

Does testing help to improve levels of attainment?

Are they effective in holding schools accountable for their performance?

How effective are performance measures such as value-added scores for schools?

Are league tables based on test results an accurate reflection of how well schools are performing

22. The rationale for national testing is that it raises standards and the initial rise in test scores for one or two years after the introduction of testing is taken as supporting this claim. However, there is no evidence that this implies real change in standards of achievement, but only the effect of greater familiarity with tests and of teaching that is strongly influenced by what is tested.¹⁹ Experience in England since the introduction of national testing mirrors that in the USA and other countries, where initial increase in scores is followed by a levelling off and in some cases a decrease in scores.²⁰ also showed that when familiar tests are replaced by alternative but equivalent ones, scores fall back to their original level. **The superficial learning that high stakes testing encourages does not generalise to new items**. Initially they may have had a positive effect in raising the amount and quality of science teaching in schools where it was weak and raising the profile of the subject but it appears that it is being replaced by a plateau where current practice, geared to current assessments is not going to improve in most schools and innovation is considered too risky.
23. Value-added measures can help but the definition of what is valuable, as determined by such measures, is narrow and may not remedy the situation. When used effectively such scores can have some benefit and act as an incentive for teachers to focus on moving some pupils to L5 or from L2-3 in Y6, for example, rather than focusing exclusively on L4. However, value-added measures in KS2 is not based solely on performance in science

¹⁷ Harlen, W. and Deakin Crick, R. (2003) Testing and motivation for learning, *Assessment in Education*, 10 (2): 169-208

¹⁸ Neill, S. R. (2002) *National Curriculum Tests: a Survey Analysed for the National Union of Teachers*, Leadership, Policy and Development Unit, University of Warwick

¹⁹ Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4):477-494

²⁰ Linn, R.L. (2000) Assessments and Accountability, *Educational Researcher*, 29 (2):4-16

which distorts the outcomes at the subject level. Contextual value-added approaches provide a better measure of the impact of a school on its pupils than raw attainment data alone. As schools realize the significance of these in the making of judgements by Ofsted, and if these became more prominent (i.e. the main measure) in performance tables, parents and others could make their own, much better informed, judgements about the effectiveness of schools

To what extent is there 'teaching to the test'

How much of a factor is 'hot-housing' in the fall-off in pupil performance from Year 6 to Year 7?

Does the importance given to test results mean that teaching generally is narrowly focused

24. When aggregate test results for groups of pupils – classes or year groups in a school – are used for setting goals and evaluating teachers and schools, thus making them 'high stakes' tests, teaching to the test is encouraged and excessive practising for tests becomes common place. Given the narrowness of what can be included in tests meant for a 'mass market', this has serious implications for the learning experiences of pupils, a claim is widely supported by research evidence^{21 22 23 24} and by Ofsted reports
25. Teaching to the test appears to be widespread, especially in Y6, and, as stated previously (see paragraph 20), success of pupils in KS2 tests in science can mask underlying problems which affect the experiences of pupils at crucial stages in their education, for example transition from primary to secondary school. With reference to science there is evidence that some pupils have already lost interest during year 6 when test preparation dominates. Intensive revision of all the content makes it harder to find something fresh to start the next key stage. Pupils are also switched off science if they have been denied the fun and practical elements of it, drilled and tested, then taught very little of the subject from May onwards in order to catch up with other aspects of the curriculum and take part in some enjoyable activities. It is also very de-motivating for pupils, who have worked hard for tests and got L4/5, to move to a secondary school which takes no account of what they achieved, does not make connections with or build on the primary curriculum and may even test them again!
26. Teaching to the test also occurs in Y9, particularly in schools where KS2-3 conversion rates have been poor, however, some secondary schools continue to view GCSEs as vastly more important than KS3 tests.

What role does assessment by teachers have in teaching and learning?

27. Teachers have a central role in both formative and summative assessment. Assessment for the purpose of helping learning (formative assessment or assessment *for* learning) is obviously best carried out by those directly involved – the teacher and learners - who can gather evidence as work proceeds and use it to identify appropriate next steps, either to overcome problems or to build on what has been achieved. Those involved are in the best position because they have access to evidence from all activities relating to all learning goals. This is what is needed for valid assessment for any purpose: that the information provided reflects as far as possible all the intended outcomes of education.

²¹ Crooks, T.J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, 58: 438-481

²² Reay, D. and Wiliam, D. (1999) 'I'll be a nothing': structure, agency and the construction of identity through assessment, *British Educational Research Journal*, 25: 343-345

²³ Gipps, C. (2005) Accountability Testing and the Implications for Teacher Professionalism, in (ed.) C.A. Dwyer, *Measurement and Research in the Accountability Era*. New Jersey: Laurence Erlbaum Associates, pp. 99-111

²⁴ ARG (Assessment Reform Group) (2002) *Testing, Motivation and Learning*. Obtainable from the ARG website: www.assessment-reform-group.org and from the CPA office of the Institute of Education, University of London

28. The arguments for using evidence collected during teaching and learning also apply to assessment for the purpose of reporting and recording progress in learning (summative assessment or assessment *of* learning). Progress towards all important goals should be reported. In addition, the information has to be reliable so that those using it can depend upon it. But information that is technically of high reliability is not useful if it does not give a valid account of learning. To preserve validity, it would be possible to use teachers' judgments and indeed to use the evidence already collected and used for formative assessment. Because formative assessment is concerned with individuals and based on the teacher's knowledge of the pupil, steps would need to be taken to ensure reliability of the information if it were used for summative purposes. Good practice in quality assurance of teachers' judgments has already been built up in many primary schools in order to meet the requirement to report teachers' judgments as well as test results in the core subjects and in non-core subjects at the end of Key Stage 2. In addition, schools often choose to use levels in their annual reports to parents. Whilst there is guidance from QCA on moderation practices, moderation is not a requirement. Where it is carried out it may take the form of meetings where teachers examine evidence of different kinds and consider how they would judge it against the criteria for performance at different levels. Some schools have developed portfolios of assessed work that represent their agreed judgments of what work at different levels looks like. This can be used as a substitute for meetings for which time is often scarce. Those who do make this time for meeting, however, find that the experience leads to better understanding of the criteria and the goals of learning and supports collaborative and continuing learning by teachers.
29. When results are used to inform teachers in other schools, particularly in transition from Key Stage 2 to Key Stage 3, it is necessary for moderation to involve meetings of teachers from different schools. Such practice is, for example, being developed in Wales, where teachers from schools in clusters meet to moderate end of Key Stage 2 assessments.
30. Assessment for learning should be at the heart of everything that happens in the classroom but in the test driven climate it does not have as much importance as it should in many schools. Too many teachers/senior leaders lack confidence and so prefer to buy or create an end of year/term/unit test and make summative judgments against marks schemes rather than develop their skills in teacher assessment. **Teachers who are required to use teacher assessment rather than tests have to engage more with the assessment criteria for the whole curriculum and ways of teaching it effectively and assessing it rather than how to prepare to answer test questions.**

National Key Stage Tests: the future

*Should the system of national tests be changed?
If so, should the tests be modified or abolished?*

31. The main reasons for proposing use of teachers' assessment go beyond reliability and validity of the information. They take us back to the involvement of pupils and the use of assessment to help learning. This is learning that leads to understanding of the important broad ideas, key skills and attitudes in various subjects that are widely applicable. In addition, an important part of preparing young people for life and work in the rapidly changing society of today and tomorrow is to help them develop awareness and understanding of the process of learning – a key aspect of meta-cognition. This involves reflecting on their learning and how it takes place, itself a goal of education.
32. There is nothing new in suggesting that these competencies are important outcomes of modern education and that their development ought to be part of education from the

earliest years. Reference to them abounds in official documents and in reports and recommendations from a number of influential groups and councils. But they are not well represented in what is assessed and in the way in which pupils' achievements are assessed. **Unless attention is paid to this, they will remain only as rhetoric**

33. The more intimately assessment is interwoven with learning, being about the process as well as the content of learning, the more important it is for assessment to be in the hands of those directly involved – the teachers and pupils. As pupils become more able to take part in their own assessment they will be able to participate in assessment for summative as well as formative purposes. Research shows that involvement in formative assessment can support the self-esteem and promote the further learning of pupils^{25 26} and there is every reason for confidence that involvement in summative assessment will have the same impact. It is far easier to see this as a possibility when the summative assessment is carried out by teachers and based on the regular work of pupils than when it is on the basis of tests or tasks that are given, unseen, solely for the purpose of grading or assigning a level

The Secretary of State has suggested that there should be a move to more personalised assessment to measure how a pupil's level of attainment has improved over time. Pilot areas to test proposals have just been announced. Would the introduction of this kind of assessment make it possible to make an overall judgment on a school's performance? Would it be possible to make meaningful comparisons between different schools? What effect would testing at different times have on pupils and schools? Would it create pressure on schools to push pupils to take tests earlier?

34. A move towards a more personalised system of assessment would be welcomed but steps must be taken to ensure that the lessons of the past need to be learnt and that the replacement arrangements do not simply recreate the high stakes regime in a different form. **ASE has concerns the recent proposals²⁷ will increase the burden of testing on pupils rather than ease it.** The resulting pressures could, in some schools, have a detrimental effect on the curriculum and on the wellbeing of pupils as well as creating a sequence of 'testing weeks' each with intensive revision periods, further reducing actual teaching time and engagement.
35. Careful consideration needs to be given to the potential unintended consequences in the medium to long term leading to increased pressure on schools/departments to get more "points" by getting more children to reach specific levels in SATs and to do more GCSEs/AS and A-levels. The social impact of children who have been moved out of their age groups should be a key factor in decisions regarding the time at which children should take tests. The development of the "whole child" should be seen as more important than 'points accumulation' but the degree to which this would indeed take priority might vary from school to school depending on their objectives.

If Key Stage tests remain, what should they be seeking to measure? If, for example, performance at Level 4 is the average level of attainment for an eleven year old, what proportion of children is it reasonable to expect to achieve at or above that level? How are the different levels of performance expected at each age decided on? Is there broad agreement that the levels are appropriate and meaningful?

36. Any continuation of key stage tests should include more formative assessment, including oral and practical activities, to really get at what specific knowledge, understanding and

²⁵ Black, P. & Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education*, 5: 1-74

²⁶ Wiliam, D., Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement, *Assessment in Education*, 11 (1) 49-65

²⁷ DfES (2007) *Making Good Progress* Consultation. London: Department for Education and Skills

skills pupils actually have. If comparisons are to be drawn then it is essential that the criteria on which they are made must be openly agreed, set and applied. Expectations regarding performance at particular levels is about the establishment of a clear understanding by **all** about what a particular level "looks like". It is unhelpful, for example, that L4+ are **changing** for KS3 but not for KS2 hence potentially fuelling a debate about when "Level 4 is not a Level 4".

Testing and assessment at 16 and after

Is the testing and assessment in "summative" tests (for example, GCSE, AS, A2) fit for purpose?

Are the changes to GCSE coursework due to come into effect in 2009 reasonable? What alternative forms of assessment might be used?

What are the benefits of exams and coursework? How should they work together?

What should the balance between them be?

Will the ways in which the new 14-19 diplomas are to be assessed impact on other qualifications, such as GCSE?

Is holding formal summative tests at ages 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?

To what extent is frequent, modular assessment altering both the scope of teaching and the style of teaching?

How does the national assessment system interact with university entrance? What does it mean for a national system of testing and assessment that universities are setting entrance tests as individual institutions

37. The testing and assessment arrangements at 16 and after suffer many of the shortcomings described for the National Testing pre-16. In particular the balance between formative and summative assessment needs to be revisited, as does the type and structure of questions that are set. Whilst there are good reasons to provide structured objective questions as part of a testing process, using them almost exclusively reduces the scope of what can be meaningfully assessed. Especially with higher level examinations the lack of opportunity to develop a full argument through an essay-type answer means there is little or no evidence to determine whether a student can indeed use their knowledge and understanding to reason and present a cogent case.
38. Despite the potential dangers of coursework there is a strong case that students should be required to produce extended reports of work they have undertaken. **Science in particular is a subject that involves investigations which require extended thought and experimentation and therefore this should have a place within the assessment process in order to recognise students' achievements in this aspect of their studies.**
39. Testing at post-16 is complicated by the plethora of qualifications that are available. In itself this is not a negative comment because a single style of qualification would be inappropriate to meet the wide range of needs. However, despite the efforts to develop a national qualifications framework, there is still much to be done to rationalise what is a very confusing situation. The introduction of the 14-19 Diplomas is in part an attempt to help this situation but as yet it is too early to say how effective they will be. One thing that is clear from everything that is happening 14-19 is the **need to reduce the burden of assessment on students and to find ways of reducing the fragmentation and interruptions to learning** that are, in part, the outcome of the type of modular system the currently exists.
40. Another consequence of a plethora of qualifications and the use of them to provide a nationwide system of testing is the problem of defining the level of difficulty for comparisons between qualifications as well as between different subjects within a particular set of qualifications. Science subjects along with mathematics, for example, are

often claimed to be more difficult than other subjects but this is disputed by, among others, QCA. Such issues need further investigation in order to address the situation.

APPENDIX ONE

The Association for Science Education

The Association for Science Education is the largest subject association in the UK, with approximately 18,000 members including teachers, technicians and others involved in science education. The Association plays a significant role in promoting excellence in teaching and learning science in schools and colleges. Working closely with the science professional bodies, industry and business, ASE provides a UK-wide network bringing together individuals and organisations to share good ideas, tackle challenges in science teaching, develop resources and foster high quality continuing professional development.

The objects and purposes of ASE are stated in its Charter of Incorporation as the promoting of education by the following means.

- *Improving the teaching of science;*
- *Providing an authoritative medium through which opinions of teachers of science may be expressed on educational matters; and*
- *Affording a means of communication among all persons and bodies of persons concerned with the teaching of science in particular and education in general.*

In a more modern context, The Association for Science Education aims to promote excellence in science teaching and learning by:

- (a) Encouraging participation in science education and increasing both new membership and the retention of existing members.
- (b) Enhancing professionalism for teachers, technicians and others through provision of high quality continuing professional development and promotion of chartered status.
- (c) Working in partnership with other organisations, thus maintaining and strengthening its position in influencing policy and its reputation for delivering cutting edge initiatives for its members and, through them, to the wider science education community.

Further details of the ASE and its regional, national and international activities can be found on its web-site (www.ase.org.uk).

The Association of Tutors in Science Education (ATSE)

This Special Interest Group of The Association for Science Education exists to:

- further the aims of The Association for Science Education;
- support the work of Science Tutors, mentors and others working in initial teacher education throughout the UK;
- facilitate the exchange of ideas about science education, and alert national agencies to issues of concern to the membership.

National Advisers and Inspectors Group for Science (NAIGS)

This Special Interest Group of The Association for Science Education exists to:

- further the aims of The Association for Science Education;
- support the work of Science advisers, inspectors and others working in a science advisory or support capacity throughout the UK;
- facilitate the exchange of ideas about science education, and alert national agencies to issues of concern to the membership.

The SCORE Partnership

The SCORE partnership aims to bring collective action and a strategic approach to strengthening science education, and believes that the key to maximising the impact of its efforts, especially their influence on government, lies in a greater degree of collaboration and in having a sense of common purpose. Through this collective action, the partnership aims to increase its influence over the direction of science education in the years to come, in particular over teacher supply and retention, curriculum development, assessment, delivery of support to teachers and students, and strategies for reaching all young people regardless of age, background, level of ability, gender, ethnic origin and geographical location.

Association for Science Education www.ase.org.uk

Biosciences Federation www.bsf.ac.uk

Institute of Biology www.iob.org

Institute of Physics www.iop.org

Royal Society www.royalsoc.ac.uk

Royal Society of Chemistry www.rsc.org

Science Council www.sciencecouncil.org