

National Curriculum Task Group on Assessment and Testing

A Report

Department of Education and Science and the Welsh Office

REPORT OF THE TASK GROUP ON ASSESSMENT AND TESTING

CONTENTS

Preface and Glossary
Letter to the Secretary of State

Sections of Main Report:

Paragraphs:

INTRODUCTION

I.	Our starting point	1 - 6
II.	Present practice	7 - 12
III.	Concerns to be addressed	13 - 19
IV.	The structure of the report	20 - 22

PURPOSES AND PRINCIPLES

V.	Purposes	23 - 30
VI.	Profile components in the assessment framework	31 - 39
VII.	Assessment and testing methods	40 - 50
VIII.	Bias in assessments	51 - 53
IX.	Ensuring confidence in national assessment	54 - 63
X.	Moderation	64 - 90

THE ASSESSMENT SYSTEM IN PRACTICE

XI.	Progression, scales and ages	91 - 124
XII.	Reporting	125 - 138
XIII.	The primary stage	139 - 157
XIV.	The secondary stage	158 - 167
XV.	The assessment of children with special educational needs	168 - 173

IMPLEMENTATION

XVI.	The brief for the subject groups	174 - 188
XVII.	Phasing	189 - 199

A SYSTEM OF SUPPORT

XVIII.	The support structure	200 - 205
XIX.	Inset and other resource needs	206 - 214
XX.	The local education authority's role	215 - 217
XXI.	Future work for the task group	218

SUMMARY AND CONCLUSIONS

XXII.	Conclusions	219 - 226
XXIII.	A List of Recommendations	227

PREFACE AND GLOSSARY

This report of the Task Group on Assessment and Testing ventures into a specialised field with its own technical vocabulary. We have tried to avoid technical jargon wherever possible, but it has sometimes been necessary to use terms which are unfamiliar to the layman or whose meaning here is different from that of everyday usage. We explain some of these terms below. Others are explained where they first appear in the report - including those we have specially coined to label ideas newly devised by the Group

Assessment

A general term encompassing all methods customarily used to appraise performance of an individual pupil or a group. It may refer to a broad appraisal including many sources of evidence and many aspects of a pupil's knowledge, understanding, skills and attitudes; or to a particular occasion or instrument. An assessment instrument may be any method or procedure, formal or informal, for producing information about pupils: e.g. a written test paper, an interview schedule, a measurement task using equipment, a class quiz.

Test

Strictly, any assessment conducted within formal and specified procedures, designed to ensure comparability of results between different test administrators and between different test occasions. For some it implies a set of written questions, externally prescribed, with written responses marked according to rigid rules; for others, any of a broad range of assessment instruments with standardised rules of administration and marking which ensure comparability of results. This report uses the term in the latter, broader sense. Tests intended for widespread use are commonly pilot-tested with a representative national sample. A *standardised test* is one published with marking instructions and data on the levels of response established for a national population sample, at a given age, or with different data for several ages.

Tasks and projects

A task is any activity in which a pupil takes part and which could be used for assessment. Examples are a written test, an essay, a practical job such as designing or constructing an artefact, or conducting an experiment. Tasks which are open-ended and extensive in scope and/or in time are often called projects: these may require independent planning and/or research on the part of the pupil.

Item bank

An orderly collection of test questions or tasks, so assembled and labelled that users may select from it sets of tasks for particular assessment purposes. Those in charge of such a bank usually exert quality control on the questions put in it.

Ranking, rating and marking

Ranking involves putting pupils' performances in order of relative quality. The set of rankings forms a scale. Where a scale simply has some order, it may be called a ranking scale. If the order is related to given criteria, it may be called a rating scale. Different performances are often assigned numerical marks and the marks added. The numbers may then be treated like measurements of length or weight, as if, for example, 50 marks represented twice the performance of 25 marks. Mark scales can be misleading in this respect.

Norm-referencing

An assessment system in which pupils are placed in rank order and pre-determined proportions are placed in the various grades. It implies that the grade given to a particular pupil depends upon a comparison between that pupil's performance and those of all the other pupils in the group, rather than upon the absolute quality of the performance.

Criterion-referencing

As defined for this report, an assessment system in which an award or grade is made on the basis of the quality of the performance of a pupil irrespective of the performance of other pupils: this implies that teachers and pupils be given clear descriptions of the performances being sought. This is a broader and less exacting definition than that used by some authors.

Aggregation

Combining a pupil's marks or grades on a set of different assessment items to give a single mark or grade. This may be done in different ways: for example, two numerical marks may be added together, or two grades may be combined according to rules (e.g. AA, AB or BA all give a complete A).

Level

In this report, *level* indicates achievement of a particular set of performance criteria which have been defined within a "profile component" (see section VI), or cluster of attainment targets. In general, a higher level will imply achievement of greater sophistication or range than that for a lower level.

Grades

Different qualities of response to a common assessment demand or task. "Pass/fail" examinations are usually referred to as ungraded. Criterion-referenced assessments are "pass/fail" insofar as pupils' performances are judged with respect to unitary criteria. The equivalents to grades in a criterion-referenced system are results summed over the attainment target criteria, so that success is judged relative to those attainment targets which have been met. In this report we distinguish

between the quality of response to the same attainment targets (*grades*), and the degrees of success in reaching a collection of attainment targets (*levels*).

Moderation

The process of checking comparability of different assessors' judgements of different groups of pupils.

The term can also refer to a procedure for determining a single set of marks or ratings when two sets are available for the same pupils on the same profile components, e.g. one set from a teacher's course assessments and one from a national test. One set (usually the teachers') may be adjusted to give the same mean and spread as the other. This is a statistical operation usually called scaling. There are other forms of statistical moderation: for example, when the "results" are expressed as grouped percentage distributions of pupils, adjustments can involve deciding which borderline pupils should change their group status. This can be a more sensitive procedure than scaling.

Calibration

Where procedures for equating results of the assessments are statistical, the procedure is sometimes called calibration, to distinguish from procedures in which the judgement of assessors is involved (moderation). This scaling is a form of calibration. Procedures using both statistical data and judgements are sometimes called "intelligent statistical moderation".

Skill

Ability to perform, usually acquired by training: can be displayed in many types of performance, e.g. manual, craft, design, communication, observation, measurement.

Concept

A generalised idea or notion; identifying something as an example of a particular concept enables one to infer how to treat it.

Knowledge

That which is known in an ordered way and can be used; in this sense different from just a collection of unrelated information.

Understanding

The ability to select and use knowledge and skills over a variety of contexts to meet a variety of demands.



King's College London (KQC)

UNIVERSITY OF LONDON

CHELSEA CAMPUS
552 KING'S ROAD, LONDON SW10 0UA TELEPHONE: 01-351 2488

CENTRE FOR EDUCATIONAL STUDIES

HEAD OF DEPARTMENT
Professor P. J. Black, OBE, KSG

24 December 1987

Rt Hon Kenneth Baker MP
Secretary of State for Education and Science
Department of Education and Science
Elizabeth House
York Road
London SE1 7PH

Dear Secretary of State

TASK GROUP ON ASSESSMENT AND TESTING

I present herewith the main report of the Task Group which you set up last July to advise on assessment and testing within the national curriculum.

Because of the short time available we have not been able to give adequate attention to one part of our remit: we are not yet in a position to make recommendations about the various services and arrangements required to support the system that we propose. We discuss some of the issues involved in the enclosed report, but we would like to give more considered advice about these in about two months' time. We might also wish at that stage to give fuller elaboration on other issues, as envisaged in your letter of guidance sent to us last October.

I am happy to assure you that this report has the support of all members of our group. I am grateful to many members of your Department for the help and support that they have given during our work.

Yours sincerely

Professor P.J. Black
Chairman
Task Group on Assessment and Testing

REPORT OF THE TASK GROUP ON ASSESSMENT AND TESTING

INTRODUCTION

1. Our starting point

1. The Task Group has been asked to advise on the practical considerations governing assessment within the national curriculum. We start from the following propositions.
2. A school can function effectively only if it has adopted:
 - clear aims and objectives;
 - ways of gauging the achievement of these;
 - comprehensible language for communicating the extent of those achievements to pupils, their parents and teachers, and to the wider community, so that everyone involved can take informed decisions about future action.
3. Promoting children's learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set, and pupils' progress charted and expressed. It can yield a basis for planning the next educational steps in response to children's needs. By facilitating dialogue between teachers, it can enhance professional skills and help the school as a whole to strengthen learning across the curriculum and throughout its age range.
4. The assessment process itself should not determine what is to be taught and learned. It should be the servant, not the master, of the curriculum. Yet it should not simply be a bolt-on addition at the end. Rather, it should be an integral part of the educational process, continually providing both "feedback" and "feedforward". It therefore needs to be incorporated systematically into teaching strategies and practices at all levels. Since the results of assessment can serve a number of different purposes, these purposes have to be kept in mind when the arrangements for assessment are designed.
5. Any system of assessment should satisfy certain general criteria. For the purpose of national assessment we give priority to the following four criteria:
 - the assessment results should give direct information about pupils' achievement in relation to objectives: they should be **criterion-referenced**;
 - the results should provide a basis for decisions about pupils' further learning needs: they should be **formative**;
 - the scales or grades should be capable of comparison across classes and schools, if teachers, pupils and parents are to share a common language and common standards: so the assessments should be calibrated or **moderated**;
 - the ways in which criteria and scales are set up and used should relate to expected

routes of educational development, giving some continuity to a pupil's assessment at different ages: the assessments should relate to progression.

6. Our recommendations have been composed in the light of these. Unless the criteria are met, the potential value of national assessment in assisting learning and supporting the professional development of teachers is unlikely to be realised.

II. Present practice

Assessment in schools

7. Generally speaking, teachers and pupils are well aware of the importance of assessment. A great deal of effective assessment is already carried out, often in an informal manner within the individual classroom. Many schools have established common systems for use throughout the school. In primary schools these systems typically require agreement and joint action by all the teachers in the school, while in secondary schools collaboration is focused within subject departments to establish common methods and criteria. Where this happens the results become one of the main inputs in a review of the work of a school or a department. Elsewhere, common guidelines may be used throughout a school. A shift from norm-referencing to criterion-referencing improves the value of this. Systematic records of pupils' progress can then be helpful in the transition between year groups and between schools. Records become the basis for discussion with pupils, and with their parents. For all of these uses assessment which provides details of achievement in different aspects of any subject is more valuable than a single overall grade.
8. Teachers' own use and interpretation of assessment and testing data is, however, limited by the range of instruments available, by the lack of time and resources required to develop new instruments, and by the lack of external reference and a basis for comparison. External agencies involved with teachers in developing the provision and use of new assessment and testing procedures – for example through the Assessment of Performance Unit, the various new Graded Assessment schemes, or in new GCSE work – have found that both primary and secondary teachers welcome the professional development flowing from better assessment methods and exchange of information about expected outcomes. Experience, particularly but not exclusively in mathematics, suggests that paying greater attention to assessment, with improved means for assessing progress in relation to curriculum targets, helps teachers to learn important lessons about their pupils. This often reveals that the needs of the less able, or the competence of the most able, have hitherto been under-estimated.

External initiatives

9. Many helpful developments in assessment have arisen from initiatives outside the individual school, including those of local education authorities, the Department of Education and Science, the Secondary Examinations Council and the new examining groups. These show that initiatives at a national level which lead to production and development of assessment instruments, and marking and moderation procedures, can be of great value within schools if carefully considered and sensitively deployed.

10. Large scale use of external standardised tests is not a new or unusual practice. In the primary phase, almost all children take such tests at some time (see Appendix C). Our evidence has shown however that these externally provided measures are not usually related, in either construction or subsequent use, to the normal work of teachers. Classroom assessments are hardly affected by them. In addition, whilst some of the more recent tests reflect good assessment practice, many older but still widely used tests are narrow in scope and employ a limited set of techniques. Although external tests have been of some value in screening to identify children with special learning needs, they have rarely been used to assist normal teaching work. The resources devoted to such testing have been scarcely justified, and the more limited types of test have themselves given testing in general a reputation for narrowness and irrelevance.

Local authority systems

11. Many local authorities have developed guidelines, and even common instruments and systems, for assessment in their schools. Some require use of standardised tests by all schools and treat the results as performance indicators. The difficulty here is commonly the lack of relationship between these tests and the learning aims actually pursued, so that teachers often regard the tests as time-wasting, or even oppressive interference. Other authorities provide only very general guidelines, believing that individual teachers must take full responsibility for such work. The difficulty here is that little external help can be provided. There is therefore no shared language and no fostering of common perceptions and criteria to help teachers and schools to improve their work, or to help parents and others to understand the performance of a school. Authorities rarely seem to have achieved an effectively integrated system in which centrally provided assessment resources give useful overall information, and also help by indicating appropriate methods and providing calibration, which teachers can make use of in improving their own work in their own ways. Several, however, are seeking to develop systems with these characteristics.

Other countries

12. Many other countries have a national curriculum or uniform curricular guidelines for each region. Many have examinations at the end of compulsory schooling, and for selective access to further education, but only a few have national assessment schemes for younger ages. In most cases, assessments at these younger ages are left to teachers. In almost all European countries the regular assessment by teachers, using mainly written work which they set and mark, is designed to indicate pupils' progress in relation to a national or regional curriculum. Results of internal assessments are sometimes communicated to parents, according to national or local law, in terms of grades on a specified scale (1 to 5 in Sweden with 5 being high, 1 to 6 in West Germany with 6 being low). In West Germany however these grades are only loosely defined and there is no procedure for moderating assessment or for calibration between schools. In Sweden standardised achievement tests are used for calibration; these are compulsory beyond 14, voluntary – but widely used – from about age 9 to 14. Publication of test results for use outside the school is virtually non-existent, except in some States in the USA. These results have been published with adjustments for socio-economic background effects: but while this can give a fairer reflection of the achievements of a school, it has made the information too complicated for the general public to understand. No country appears to have a national assessment system which is well developed in relation to **formative** purposes and to a framework of **progression**. It is worth noting also that assessment instruments developed by the APU in

this country for monitoring purposes are internationally recognised for being both innovative and of the highest quality.

III. Concerns to be addressed

13. Whilst present practices, both in this country and in others, give many examples of positive uses of assessment, no system has yet been constructed that meets all the criteria of **progression, moderation, formative** and **criterion-referenced** assessment set out in paragraph 5 above. Our task has therefore been to seek to devise such a system afresh. We believe that the model of assessment put forward in this report builds on some existing good practice and represents an advance on assessment practices in other countries. But we could not approach this task without also recognising that many are deeply opposed to any system of national assessment and testing. They fear that it may do serious damage and they see no possible advantages. In forming our recommendations, we have tried wherever possible to address these fears, in ways we summarise below.

Effects on pupils

14. Some of the fears relate to **damage to pupils**. Yet when assessment and testing are carefully aligned to the curriculum, as in Graded Assessment Schemes, one of the outstanding benefits that teachers report is the enhanced motivation of pupils. Many pupils seem to need short-term objectives which are clear and which they perceive to be attainable. The possibility of achieving these builds enthusiasm for progress in learning. Of course pupils do not always succeed, but poor performance against a target should not be seen as a prediction of personal inadequacy; it should usually be regarded as an indication of needs. Such needs must be met, perhaps first by the use of more detailed diagnostic assessments conducted usually by the school to identify specific difficulties, and then by further teaching help. It would be absurd if diagnosis of difficulty – and recognition by a pupil and the parents and teachers of the need for help – were to be avoided because of fear of failure. It would equally be wrong if such evidence about an individual were to be reported to others beyond the parents, except in confidence to those in a position to provide help. Our proposals reflect these considerations and, we believe, can positively assist children's development.

Relations between parents and schools

15. Other fears relate to possible **damage to the relationships between parents and schools**. But both parents and pupils need more intelligible information than is provided at present, particularly in the secondary phase where a parent may receive reports from several teachers, each using different scales and language to record progress. Parents are interested above all in where their child stands in terms of the general progress expected at that age. At present the usual comparison, if any, is with the class or year group within the school. A national curriculum leading to national criteria and calibration could change this, and could help schools to give firmer information and where appropriate, reassurance to parents. Similarly, employers and further education institutions, who often have to select in advance of GCSE results, cannot readily relate the reports of different schools because they cannot compare the standards applied by those schools and the different scales and levels of detail used in reports. Our proposals aim to promote and support an informed

relationship between schools, parents and others.

Effects on teachers' work

16. A more widely-voiced fear is that external tests will impose **arbitrary restrictions on teachers' own work**, and so limit and devalue their professional role. Our first response to this has been to recommend procedures in which:
- teachers' assessments over time and in normal learning contexts play an important part;
 - externally-provided methods and procedures are broad in scope and related to the curriculum attainment targets which all will share;
 - these assessment methods may often be incorporated into normal classroom activities;
 - the administration, marking and moderation procedures rely on the professional skills and mutual support of teachers, giving them both key responsibilities and communal safeguards against idiosyncrasy.
17. But we also recognise that many teachers, at all levels, will need help to incorporate any new and systematic assessment procedures into their normal classroom work. Experiences both in assessing GCSE subject coursework of pupils at ages 15 and 16 and in in-service training for primary teachers working with 7 year olds on project tasks have shown that teachers need time and support to help them adapt. They have also shown that teachers welcome the professional growth they gain from such work if they can come to share the assessment aims and translate the methods into their own style of work. These considerations indicate that provision of carefully planned and pilot-tested assessment instruments, and of in-service training directed to initiating and supporting new types of classroom work, will be essential.

Publication

18. Finally, there is a fear that results will be published in **league tables of scores**, leading to ill-informed and unfair comparisons between schools. We believe that most teachers and schools would not object to assessment results being reported to those who know the school and can interpret them in the light of a broader picture of its work and circumstances. They would object, however, to the publication of partial information which is not set in that context and is therefore potentially misleading – particularly where significant decisions are then based upon that information. It needs to be recognised that many of the outcomes of a school's work are influenced by external circumstances, notably the social and economic background of pupils' families, over which the school has no control. These factors do not of themselves affect the validity of a criterion-referenced attainment target. But judgements about the quality of a school should not be confined to the extent to which the targets are actually reached. They should also take into account the educational value added – that is, the progress it might have been reasonable to expect a school in such circumstances to secure among its pupils, bearing in mind the pattern of attainment at intake and variations in children's rates of development. We recognise that these factors are difficult to measure and interpret; this strengthens the need for information about a school's environment. We therefore consider that the publication of assessment results for any school should take

place if, and only if, this is done in the context of reports about that school as a whole, so that it can be fair to that school's work and take account so far as possible of socio-economic and other influences.

Our proposals

19. No new scheme could immediately set all such fears entirely at rest. We emphasise, however, that many of them arise from assumptions, based on experience of previous schemes and practices, which do not apply to our proposals. Our proposals are both evolutionary and radical. They are evolutionary in that they build on examples of good practice that we have studied, some of them of recent origin. They are radical in that they are developed and will be adopted on a national scale, and because our remit can be fulfilled in a positively helpful way only by some quite new departures. A more detailed discussion of this important issue is given in Section XXII.

IV. The structure of the report

20. With those considerations in mind, the following sections of this report set out the arguments and our recommendations for a national system of assessment and testing. Section V starts by discussing the purposes of the assessment system. This introduces a sequence of sections VI - X exploring the ways in which any particular performance component, representing a linked cluster of attainment targets within a subject, should be assessed and moderated. The next sequence XI - XV considers how these components may be combined to form overall assessments for pupils at and across various ages, how they should be reported, and the implications for primary and secondary schools and special educational needs. This is followed by sections XVI and XVII which set out various consequences for the tasks of the national curriculum subject groups and for phasing the introduction of national assessment working arrangements. After this, we turn in sections XVIII - XXI to implications for the setting up and support of the national system. Finally in sections XXII - XXIII we discuss some general conclusions and summarise our recommendations.
21. The report is accompanied by appendices. Appendices A, Band C respectively give the Group's terms of reference and membership, the Secretary of State's letter of guidance, and a list of meetings held. Subsequent appendices expand on issues which, for brevity's sake, have to be presented in the main report without full supporting argument or illustration.
22. We draw attention also to the need for clear terminology. The preface to this report gives the necessary specifications and these have been followed throughout; in many cases common usage is more flexible, but for our purposes this could have given rise to misleading ambiguity.

PURPOSES AND PRINCIPLES

V. Purposes

23. The terms of reference and the letter of guidance from the Secretary of State require that information derived from assessments (including tests) shall be capable of serving several purposes:
- **formative**, so that the positive achievements of a pupil may be recognised and discussed and the appropriate next steps may be planned;
 - **diagnostic**, through which learning difficulties may be scrutinised and classified so that appropriate remedial help and guidance can be provided;
 - **summative**, for the recording of the overall achievement of a pupil in a systematic way;
 - **evaluative**, by means of which some aspects of the work of a school, an LEA or other discrete part of the educational service can be assessed and/or reported upon.
24. Our attention has also been drawn to the effects all forms of assessment have on teaching and learning. We are required to propose a national assessment system which enhances teaching and learning without any increase in "the calls on teachers' and pupils' time for activities which do not directly promote learning". We judge that it would be impossible to keep the burdens on teachers and pupils within reasonable bounds if different batteries of assessments and tests had to be created to serve each separate purpose. We must therefore reflect the priority given in our brief to the need "to show what a pupil has learned and mastered".

Serving several purposes

25. Some purposes may, however, be served by combining in various ways the findings of assessments designed primarily for a different purpose. It is possible to build up a comprehensive picture of the overall achievements of a pupil by aggregating, in a structured way, the separate results of a set of assessments designed to serve formative purposes. However, if assessments were designed only for summative purposes, then formative information could not be obtained, since the summative assessments occur at the end of a phase of learning and make no attempt at throwing light on the educational history of the pupil. It is realistic to envisage, for the purpose of evaluation, ways of aggregating the information on individual pupils into accounts of the success of a school, or LEA, in facilitating the learning and achievements of those for whom it is responsible; again, the reverse process is an impossibility.

Formative and summative

26. We judge therefore that an assessment system designed for formative purposes can meet all the needs of national assessment at ages before 16. At age 16 the focus shifts from formative to summative. The GCSE is a terminal examination in the sense that it comes at the end of the compulsory period of education and thus cannot avoid summative judgements. Similar considerations would apply to other forms of assessment at that stage. Summative judgements or aggregations may be made at ages other than 16, but only at 16 does it seem appropriate for assessment components to be designed specifically for summative purposes.

Formative and diagnostic

27. We do not see the boundary between the formative and diagnostic purposes as being sharp or clear. If an assessment designed in formative terms is well matched to the pupil, it is likely to provide some information which will help in the diagnosis of strengths and weaknesses. Some assessments of this kind will be of value as indicators of learning problems which require further investigation. Further diagnostic investigations would, however, often involve the use of highly specialised tests of relevance to only a limited set of circumstances and pupils. It follows that the results of such tests should not be included in aggregated results for groups of pupils. Furthermore, a detailed diagnostic report on every pupil would lead to an excess of information and hence be counter-productive. The formative information about pupils will be passed on to their next teacher; if diagnostic reports are confined to problems outside the normal range, these will be highlighted and made more salient to the receiving teacher.

We recommend that the basis of the national assessment system be essentially formative, but designed also to indicate where there is need for more detailed diagnostic assessment. At age 16, however, it should incorporate assessment with summative functions.

Overall results for a pupil

28. The complete range of formative and diagnostic purposes will be served only if the results for each pupil are aggregated in various ways: for example, across cognate elements in part of a subject or field of study or across the whole of that subject or field. There may also be value in comparisons which place the performances of the pupil in the contexts of his or her earlier achievements and the achievements of the class as a whole. Out of these aggregations and comparisons can come judgements to be presented to the pupil and/or parent so that they are involved in making decisions about the pupil's future educational programme. Such judgements raise questions of the confidentiality of information about individuals. We see confidentiality as crucial if assessment is to serve a positive role.

We recommend that all assessment information about an individual be treated as confidential and thus confined to those who need to know in order to help that pupil.

Evaluation

29. The other purposes listed in paragraph 23 can be fulfilled by shifting the focus from individual to class, or to a group of classes in a school, or to the school, or to a group of schools within an LEA. The same processing procedures of aggregation and comparison, together with an analysis of the spread of the results around the average, can be applied to yield information of value to a range of audiences. National assessment designed to have formative/diagnostic functions and to serve as a foundation for all the purposes will give results which have a Common basis in the progress of the pupils in relation to national targets.

We recommend however that for summative and evaluative purposes results should be aggregated across classes or schools so that no individual performances can be separated out.

Assessment of attitudes

30. We draw attention to the fact that Records of Achievement serve purposes which overlap but are not identical with those discussed above. In our judgement it is appropriate for such records to include information about the attitudes of pupils, since statements of this kind will have been discussed with the pupil and inserted by agreement into a comprehensive record which is the property of, and confidential to, the pupil. But because of their wider role, the assessments associated with the attainment targets of the national curriculum need not, and should not, be subject to quite such close confidentiality as may be appropriate for Records of Achievement as a whole or for the personal components which are included in such records. We recognise the importance of educational aims which relate to the development of personal attitudes, and so accept that these might be considered by subject working groups in framing attainment targets. We judge it to be wrong, however, to *require* any part of the national system of assessment, including reporting, to be concerned with this part of the educational programme.

We therefore recommend that assessment of attitudes should not form a prescribed part of the national assessment system.

VI. Profile components in the assessment framework

31. The basic unit of information in any assessment system is the response of an individual pupil to a particular demand or task: for example, a written response to a test item, an observed action in a practical task, or a sequence of actions assessed by a teacher's rating. To base an interpretation upon only one unit of information is neither sufficient nor desirable. Any particular unit of information will depend heavily on the context in which the information is gathered. Before any such response can be interpreted with any certainty, it must be compared with the responses of the same pupil to other tasks and perhaps with the responses of other pupils to the same tasks. The standard method of making these comparisons is by aggregating the responses.

Feeding forward

32. The formative and summative purposes of assessment, concerned with the immediate needs and achievements of the pupil, necessarily involve aggregations of the responses to a range of tasks by individual pupils. As we have already said, the formative purpose is fundamental to the national assessment system. In order to serve that purpose effectively, the results must both feed back to the pupil and feed forward to the next teacher or institution. The emphasis on the formative function serves to underline the "feed forward" aspect. This is enhanced by the choice of reporting ages, which carries with it the implications that the teachers responsible for making the assessments will not be those responsible for acting upon them.

33. This feed forward role is diminished if the assessment of a pupil's progress in a subject is designed to produce only a single overall score. For example, in developing graded assessments in science, the London Group chose to assess and report separately on three components – skills (mainly practical), knowledge and understanding, and exploration (project) work. The results in these three components form a profile which represents the pupil's attainment in science. Similarly, many schemes for English recognise the need to attend separately to writing, oracy, reading comprehension, and listening: again, results on these four would give a profile. Information of value to pupils, parents and others would be lost if (say) a teacher were to add the results of writing to those of oracy, and to ignore those separate results when discussing pupils' future work. If subject attainments are presented as a profile of separately determined scores, then receiving teachers will be better able to use the information to shape their teaching strategies.

We recommend that, to realise the formative purpose of the national assessment system, pupil results in a subject should be presented as an attainment profile of the kind discussed above.

Profile components

34. The profile will thus be composed of a set of profile components. These components should be broadly based so as to be informative to the teachers of a range of subjects being studied by the pupil. In addition, each reporting age may herald the introduction of new subject areas. For these in particular, teachers will want a report on subjects already studied which goes beyond subject-specific knowledge and skills to the general skills, knowledge and understandings upon which they can build. Including in a subject profile those components which have applicability beyond that subject will serve to emphasise the underlying unity of a subject-based curriculum.

35. Defining the profile components will be the responsibility of the relevant subject working groups. We envisage that each component will comprise a cluster of attainment targets which have some homogeneity in relation to the skills, knowledge and understanding which the subject promotes. But their internal consistency should not be the only determining factor. To demand high correlation between scores on all of the attainment targets within each profile component might lead either to an unmanageable number of profile components, or (worse) to the omission on statistical grounds of educationally important

targets. There is a balance to be struck. On the one hand, a large number of components would daunt teachers and confuse parents. On the other, a small number would mean that each component might be too heterogeneous to be meaningful.

We recommend that an individual subject should report a small number (preferably no more than four and never more than six) of profile components reflecting the variety of knowledge, skills and understanding to which the subject gives rise. Wherever possible, one or more components should have more general application across the curriculum: for these a single common specification should be adopted in each of the subjects concerned.

36. Each profile component ought to be assessed separately and separately moderated. This has implications for the amount of assessment and moderation overall. We discuss in section XIII below the danger of overburdening primary school teachers in particular.

Feeding back

37. While the above argument for profile components rests on the feed forward aspect, profiles also provide feedback. The successes of the graded assessment movement point to the educational value of detailed feedback to pupils. Attainment targets, shared by teachers and pupils, appropriate to the current stage of the pupil's achievement and pointing the way forward to further progress, appear to play an important part in raising standards of achievement. Thus formative assessment must be in terms to which pupils can respond and which clearly indicate what has been and what remains to be achieved. The profile components should give meaningful information for this purpose and help to convey and draw upon the diversity of performance that a pupil may exhibit in any one subject.

Aggregation for evaluation

38. These arrangements should serve the evaluative purposes of national assessment equally well. Aggregating pupil performances to the level of the class, the school, the LEA, or the age group should help to monitor the delivery of the national curriculum. These aggregates can be based on any unit of analysis. Aggregates of performances in test items, standard assessments, attainment targets, profile components, whole subjects or larger curriculum segments all have their evaluative uses.
39. To sum up so far we envisage that the outcomes of the national assessment system will be presented as the attainments in a set of profile components some cross-curricular, some subject specific. The profile will become the main vehicle of communication in helping both pupils and teachers to evaluate their work.

VII. Assessment and testing methods

40. The process of education has the aim of changing people and their behaviour by imparting skills, increasing knowledge and understanding, and developing competences. This is a dynamic process in which individuals are constantly developing. It is important for teachers

to be aware of the levels reached by pupils, and their rate of development, in order to facilitate learning.

Scope of assessment

41. As a natural part of teaching, therefore, teachers are constantly assessing pupils to determine their progress and to plan the next stage of their learning. In the widest sense, such assessment involves a continuous comprehensive examination of all aspects of the pupil's learning, drawing on a wide variety of evidence from many sources to arrive at a general picture. This process is not sharply focussed, yields a diversity of information and tends to be only loosely related to purpose. Consequently it may not be clear how it relates to particular decisions about a pupil's educational programme.
42. In a narrower sense, the term 'assessment' is used to refer to an individual component of the total assessment process or to a particular method of assessment. Hence, it encompasses all procedures used to make an estimate or appraisal of an individual's achievement. Which of the many methods of assessment may be appropriate in particular circumstances will depend on the purpose of the assessment and the decisions to be taken in the light of the results obtained.
43. It will be clear that there are many sources of information for assessment, each of which is more or less suited to different purposes. Some of these sources are:
- general impressions;
 - marking coursework;
 - marking assignments;
 - pupils' self-assessment;
 - rating scales;
 - checklists;
 - practical tests;
 - written tests.

These have differing levels of formality and differing levels of standardisation. Some are appropriate only for making comparisons within a particular class or school (impressions, coursework); others are more generally applicable. Some assessment methods provide a richness of detail and a validity in context which cannot be obtained in more formal ways. However, these do not usually provide a manageable means of contrasting the performance of pupils who are being assessed by different teachers, whether in the same school or in different schools.

Standardisation

44. In order to secure comparability, an assessment component needs to be conducted according to formal procedures related to the particular aspects of an individual's knowledge, skills and competences being examined. A standardised assessment of this sort would be designed to elicit a particular type of response in a standardised manner, and to rate this, using standard methods, so that the pupil assessed would be assigned a score, grade or category. Standardisation allows the assessment component to be used widely by

many different teachers who, if they are appropriately trained in its use, can thereby produce results which are on a common scale. But it is usually time consuming. So if the calls on teaching time are to be contained, standard procedures can be used by teachers for only a limited proportion of a pupil's work, and the need to standardise may narrow the range of attainments being assessed.

Testing

45. At this point we have a difficulty with terminology. Our terms of reference use the words **assessment** and **testing**. The former, as explained above, has a broad inclusive meaning. The meaning given to **testing** is often too narrow. Much of the discussion about testing seems to be based on a particular image. This image is frequently of an externally-prescribed paper and pencil test, formal and unimaginative in scope, to be attempted in a set time on a formal occasion and marked according to set rules. The system that we recommend must include the use of tests, but far more broadly conceived. The aim of the system must be to provide standardised, i.e. nationally comparable, assessment results; but the range and scope of standardised assessments that can be used to perform this role is far wider than the term "test" is usually taken to imply. We emphasise this point to draw attention to the fact that our use of the word **test** will have this broader meaning. Ideally it might be better expressed by the phrase "**standard assessment task**". Both terms will be used interchangeably in this report, to specify externally provided tasks and procedures designed to produce performance data on a national scale.
46. As we have said, teachers carry out assessment using a wide range of methods, including tests. The national system that we propose will use, in combination, teachers' own assessment results together with the results they obtain with tests, i.e. with standard assessment tasks.

Variety of tasks

47. Standard assessments need not only be in written form. Indeed, the wide variety of possibilities can be explored by analysing any one task in terms of three aspects or modes, which can be defined as follows:
- the **presentation** mode - the method of delivery of the questions (oral, written, pictorial, video, computer, practical demonstration);
 - the **operation** mode - the expected method of working (mental only, written, practical, oral);
 - the **response** mode - pupils may answer in various ways (e.g. choosing one option in a multiple-choice question, writing a short prescribed response, open-ended writing, oral, practical procedure observed, practical outcome or product, computer input).
48. The art of constructing good assessment tasks is to exploit a wide range (far wider than those normally envisaged for tests) of modes of presentation, operation and response, and their numerous combinations, in order to widen the range of pupils' abilities that they reflect and so to enhance educational validity. Examples of the wide range of possibilities are set

out in Appendix E. These examples, which are all in current use or under development should serve to illustrate our view that tests, understood as standard assessment tasks, can be devised to achieve educational validity in relation to a wide range of learning aims.

49. One aspect of this expanded range of possibilities is that the discontinuity between teachers' own assessment of normal classroom work and the use by them of externally provided tests need not be a sharp one. For example, an open-ended task – presented in a prescribed way in a practical mode, requiring practical and oral work and assessed in a standard way through oral and written response by the pupil – could well be so deployed by a teacher that pupils would not necessarily be aware of any departure from normal classroom work (see Appendices D and E for examples).

Flexibility

50. Provided the requirements of national comparability are met, the system we propose allows a degree of flexibility at the school level. For example where provided externally, the particular test topics or themes may be a prescribed set, or an agency may provide a range from which schools can make their own choice. The form in which a task is administered may be prescribed, but it may not have to be done at fixed times – particularly for tests of a practical project nature which may extend over several class periods. If such flexibility is to be accommodated – which we think very desirable – then marking or grading procedures will have to be carefully standardised, and there will also have to be some means of moderation between teachers and schools (see section X).

We recommend that the national system should employ tests for which a wide range of modes of presentation, operation and response should be used so that each may be valid in relation to the attainment targets assessed. These particular tests should be called "standard assessment tasks" and they should be so designed that flexibility of form and use is allowed wherever this can be consistent with national comparability of results.

VIII. Bias in assessments

Gender and ethnic bias

51. As with most assessments, the results of the standard assessments are likely to show differences between groups. It is possible that there will be sex differences in some attainment areas; for example, on average girls score more highly than boys in certain writing tasks. Generally, though, the position is not simple – in mathematics boys do better in some topic areas and girls in others. There is also some evidence that different types of test favour one or other group (e.g. boys may do better in multiple-choice tests). In Appendix F we reproduce a paper from the Equal Opportunities Commission which discusses such problems in detail. Similarly, problems of bias may arise in relation to differences between the various ethnic groups.
52. Two procedures should, in our view, be incorporated into the processes of development

and analysis of standard assessments. Firstly, the agencies responsible for developing the standard assessments should incorporate formal reviews of proposed material for any evident sex or race stereotyping, and ensure a proper balance. This would normally be carried out in collaboration with panels of teachers. Secondly, the assessment devices should be statistically examined for evidence of item bias (a form of invalidity) and as far as possible the sources of such bias should be eliminated. Established statistical procedures exist for the detection of item bias.

We recommend that assessment tasks be reviewed regularly for evidence of bias, particularly in respect of gender and race.

Language difficulties

53. There may be difficulties presented by those whose first language is not English. Head teachers might exempt such children from tests where this problem is so severe as to render the assessment unworkable. Yet it should be recognised that to record a low level of performance for this reason would be no reflection on a pupil's general ability but merely an indication that the pupil needed special help in English language skills. Assessment in other skills and understanding, particularly at age 7, should, wherever practicable and necessary, be conducted in the pupil's first language.

IX. Ensuring confidence in national assessment

54. The national assessment system must be designed to yield reports of pupil performance in which a wide variety of users can have confidence. This confidence arises from the reliability of the observations made and the validity of the interpretations placed on them. Both reliability and validity contribute to a statistical estimate of confidence. In both cases confidence can be increased by using more, and more varied, types of test or task. Repetition of assessments is, however, a potential burden on teachers and pupils alike and can distract from teaching and learning. The national assessment system must therefore minimise the amount of information to be collected while maximising confidence in its interpretation.
55. Drawing the conclusion that a pupil has achieved the objectives of the national curriculum involves a chain with five main links; each carries the possibility of error.

The national curriculum must be clearly communicated

56. The first risk to confidence arises from lack of clarity in the definition of what has to be learned and assessed. In the past what is to be assessed has often been the only clear expression of what is to be taught and this has often led to a narrowing of the curriculum. This tendency can be reduced if the process can start from agreement about what has to be learned in terms of attainment targets. Illustration of what these mean, and imply, can then be communicated most clearly by means of specimen tasks with sample responses.

We recommend that attainment targets be exemplified as far as possible using specimen tasks. Such tasks can then assist in the communication of these targets.

57. Clarity of definition should not be confused with reduction to a few key elements. The curriculum is inherently complex and the subject working groups should be encouraged to mirror that complexity. We assume that there will be more attainment targets to be met at any reporting age than can reasonably be assessed on a single occasion for national assessment purposes. This problem affects both the choice of methods and the number of sampling occasions. These two issues are discussed separately in paragraphs 58 to 60 below.

National tests must use a variety of methods

58. Secondly, because targets may have to be sampled, this carries the danger of narrowing the scope of assessment. This danger is enhanced by the fact that some kinds of tasks are easier and cheaper to set and mark than others. Use of only one kind of assessment instrument in the interests of economy can lead to undue emphasis on some targets at the expense of others, threatening distortion of curriculum coverage. A broad range of assessment instruments sampling a broad range of attainment targets will further discourage the narrowing tendency to "teach to the test".
59. We envisage that tests in a particular subject might cover up to half a dozen profile components. The information gained in what must be a limited time can be increased if items cover (say) application of knowledge and are designed to give information both about the knowledge and about the skills of application. Written test responses might be supplemented by teacher observations of skills made in a systematic way in the same or similar externally defined contexts. Together these can provide a sample covering the entire syllabus. The assessment should include items and activities in which pupils have to supply extensive ideas and information in their responses: error analysis of such data can be very revealing of pupils' learning difficulties.

We recommend that a mixture of standardised assessment instruments including tests, practical tasks and observations be used in the national assessment system in order to minimise curriculum distortion.

Observed responses must be shown to be typical

60. The third factor is that pupils' responses to tasks are influenced not only by their understanding of what the task involves but also by how and when the task is presented. Pupils perform differently on different days and in different circumstances. Any single response is open to these risks. They are minimised in teachers' ratings of longer term performance because those are based on many occasions and many different types of task. Teachers' ratings have the further advantage of flexibility. They can be focused on the targets of the national curriculum without unduly restricting the school's choice of ways of meeting these targets. That freedom of professional decision is an important aspect of the national curriculum. Externally prescribed tests or tasks cannot sample responses over many occasions because this would require undue interference with teaching and learning.

We recommend that teachers' ratings of pupil performance should be used as a fundamental element of the national assessment system. Just as with the

national tests or tasks, teachers' own ratings should be derived from a variety of methods of evoking and assessing pupils' responses.

Performance must be interpreted in terms of achievement

61. Fourthly, the effects of context described above are compounded, and thus likely to be confused, with more fundamental differences in pupil performance. Targets for any reporting age will be achieved by different pupils in different sequences. If reliable results were to be required about every single attainment target, an excessive amount of assessment would be needed in order to disentangle the effects of context from more fundamental individual differences in the sequence of achieving targets. But profile components based on *aggregations* of target achievements *can* be reported with acceptable degrees of confidence. Thus the case for specifying only a small number of profile components in order to support clear communication is reinforced by the argument presented here that greater confidence follows from the aggregation of targets into profile components.

We recommend that when the subject working groups provide guidance on the aggregation of targets into a small number of profile components, they should have regard to the need for each component to lead to a report in which reasonable confidence is possible.

Interpretations of achievement must be related to national standards

62. Finally, national standards are directly conveyed in the mark schemes of standardised tests. If teachers' rating scales are to communicate national standards clearly, they should not only report on a small number of profile components but also use a *short* scale for each component so that each scale point can be tied to a description of a significant level of attainment. Teacher ratings are, however, open to misinterpretation, since they are strongly influenced by local expectations of achievement. On the other hand, teachers' expectations have an important role in providing feedback to help to ensure that national standards, based on subject-specific requirements, are realistic. Moderation is the key to effective national standardisation. We discuss what moderation involves in the next section.

We recommend that teachers' ratings be moderated in such a way as to convey and to inform national standards.

63. To achieve confidence in the face of these risks of error is not such a demanding requirement as at first appears. A combination of teachers' ratings and standardised assessment tasks can engender confidence in the interpretation of achievement. The combination will facilitate fair sampling of the national curriculum, coverage of different circumstances of assessment on different occasions, and a range of evidence as the basis of report. Individually the instruments are open to misinterpretation but in combination they minimise the risks.

We therefore recommend that the national assessment system is based on a combination of moderated teachers' ratings and standardised assessment tasks.

X. Moderation

Purpose of moderation

64. The process of bringing individual judgements into line with general standards is known as moderation and is an essential part of an assessment system.
65. We are all, as individuals, persuaded that those things which occur frequently in our experience are 'normal'. Schools' catchment areas are not representative of the national population but they condition teachers' expectations of what is a 'normal' average and spread of achievement. These expectations are powerfully reinforced by successive year groups of pupils. In the absence of equally powerful external evidence, teachers' expectations become the teachers' standards.
66. Teachers' expectations of individual pupils in the class create further problems, loosely referred to as the 'halo' effect. In the absence of a close definition of what to look for and how to observe it, we look for confirmation of our expectations. Research evidence shows that teachers' rank orders of pupils' performance may vary systematically from rank orders provided by test users.
67. All three statistical features of a set of school-based assessments – the mean, the spread and the rank order – are affected by fallibility in judgement, not because teachers are untrustworthy, but because they are human. The choice of a particular moderation method should recognise that distinction.

Methods of moderation

68. Methods of moderation have twin functions: to communicate general standards to individual assessors and to control aberrations from general standards by appropriate adjustments. Three methods are in common use:
 - scaling on a reference test;
 - inspection by visiting moderators; and
 - group moderation bringing teachers together to discuss their assessments.

Each differs in its relative emphasis on communication and control.

69. **A reference test** is any external instrument which has been set to the whole population of pupils who are assessed by their teachers. Scaling assumes that performance on the test is a better guide to the average and spread of the performances in a class than are the teacher's assessments and so the teacher's marks are adjusted – scaled – to bring them into line with the marks on the reference test. Usually only the rank order remains sacrosanct. The validity of the technique depends upon the reference test reflecting precisely the same features as the teacher's ratings.

70. The standardised assessment tasks of the national assessment system could be used as reference tests against which to scale teachers' ratings. This method would be economical, but there would be a price to be paid. For clear communication to teachers and parents about the attainment of the targets, the marks have to be meaningful, and mark ranges should be therefore be short. Thus it may be educationally desirable to use a reporting scale with only four levels, but any statistical adjustment then involves switching between two adjacent levels out of the four (for example, regrading all level 1 results into level 2; see paragraphs 97ff), which would usually be too gross an adjustment. On the other hand, the statistics of scaling require the teacher to use long mark ranges so that sensible timescale mark adjustments can be made. But this is in conflict with the need to communicate in terms of broad levels of performance for which long mark ranges are not necessary. Another disadvantage of scaling on a reference test is that because the test alone determines the level of reported achievement, teaching becomes dominated by test technique.
71. **Moderation by inspection** avoids this problem, but still has an overt emphasis on control, rendering communication one-sided. Furthermore, the control function itself is not efficient because it must suffer from incompleteness of data: only the final outcomes of tasks can be retained for inspection. Special arrangements may be made for the visiting moderator to observe processes and skills in operation, but these occasions are bound to be artificial and to become in effect an external practical examination.
72. In **group moderation** the emphasis is on communication through discussion and exchange of samples at a meeting. It is the favoured approach when examination boards scrutinise their own standards. Group moderation allows examiners to clarify in discussion both the objectives of their syllabuses and the bases of their value judgements. It is the only one of the three moderation methods which would enable the professional judgements of teachers to inform the development of the national curriculum. At the moment it is time-consuming and therefore expensive but it will be less so when assessment is based on explicit criteria.
73. Control can be built into group moderation. Teachers could bring to a group meeting two sets of results for each profile component. The first would be their own ratings; the second would be the results from the national tests. Both would be expressed as the distribution of their pupil groups over the levels of the national assessment scheme. A first task of a moderation group would be to examine how well the patterns of the two matched for each group of pupils. To take a simple example, if the results for a profile component were to be in terms of percentages of pupils in each of levels 1, 2 or 3, then teacher ratings might give 25% of their pupils at level 1, 50% at level 2, and 25% at level 3. On the other hand the national test results for the same group of pupils might indicate 20%, 60% and 20% respectively. The meeting would also have available the distributions over the whole group, both of the teacher ratings and of the results of the national tests (standard assessment tasks). They might also have results from all groups in an LEA; eventually, values derived from national norms may also be used, whether obtained in national trials or from results of previous years.

How group moderation works

74. The group moderation meeting would first explore any general lack of matching between the overall teacher rating distribution and the overall distribution of results on the national tests. The general aim would be to adjust the overall teacher rating results to match the overall results of the national tests; if the group were to have clear and agreed reasons for not doing this, these should be reported to the LEA and eventually to the body responsible for operating the national assessment system. Such reports should lead to a new scrutiny for possible defects in the tests if enough agreed discrepancies were to be found.

75. The meeting could then go on to explore discrepancies, in the pattern of the results of particular schools or groups, using samples of pupils' work and knowledge of the circumstances of schools. The moderation group's aim here would be to arrive at a final distribution for each school or pupil group. In general this would be the distribution on the national tests, with the implication that teachers' ratings would need adjustment, but departures from this could be approved if the group as a whole could be convinced that they were justified in particular cases. The same pattern-matching technique could be used by LEAs so that they could check that local group practice did not begin to diverge from national expectations, without the need for a cumbersome hierarchy of subordinate groups based on representatives of local groups. It should be noted that a distribution such as 20%/60%/20% over three levels is more informative than an arithmetic mean. Distributions of 20% / 60% / 20% and of 10% / 80% / 10% could give the same overall mean score but for most purposes they are significantly different from one another.

76. The controlling element in this method of moderation can be achieved without a group meeting. The value of group moderation, however, lies in communication. It affords the opportunity for teachers to discuss possible interpretations of the most frequently occurring pupil profiles and their formative implications. Mismatches between internal and external assessments can be discussed in terms of the interpretation of the national curriculum and the effectiveness or otherwise of national assessment instruments. These professional deliberations have a valuable staff development function while at the same time, with appropriate documentation of decisions, providing formative feedback to the relevant subject groups concerned with the development of the national curriculum. The national assessment system is a communication system and professional communication should be its central core. A detailed account of a moderation meeting, based on experience of such work in the GCE system, is given in Appendix H.

77. It may not be possible to undertake group moderation for all profile components every year. In that case it could be undertaken in alternate years and experience may show that it is more essential, or more professionally valuable, for some components than for others. Where moderation is not undertaken through group meetings, it should proceed by simply adjusting the distributions to agree with those of the national testing.

We recommend that group moderation be an integral part of the national assessment system. We recommend that it be used to produce the agreed combination of moderated teachers' ratings and the results of the national tests.

Supplementary assessment

78. At this stage of the assessment process, a means has to be found of resolving the minority of individual cases where the internal and external profiles do not agree, and of confirming tentative diagnoses of serious pupil difficulties.

We recommend that further assessment instruments be available for teachers to use in cases where they require additional evidence about particular pupils.

79. Such further assessment instruments would serve three purposes. Some, specially developed, should be narrowly focussed on particular learning problems to confirm tentative diagnoses. The emphasis would be on error analysis but outcomes should be related to the same scale points as teacher ratings. The other two purposes could be met without additional test construction. To resolve discrepancies, national assessment instruments from previous years could be used. For pupils whose initial performance is out of the normal range for the age group, national assessment instruments from adjacent ages could be used. Most pupils would not be involved in this phase so demands on the teacher's time should be manageable. None of the instruments used for these three purposes would be the same as the more detailed diagnostic tests needed for pupils with particularly unusual difficulties.
80. Group moderation, with follow-up observations in a sequential and 'tailored' approach to assessment, would meet the objective of sound formative reports from teachers about, and to, individual pupils and to their parents, and from teacher to teacher. The soundness of this approach would be underpinned by the professional responsibility of teachers for their individual pupils.

We recommend that the final reports on individual pupils to their parents should be the responsibility of the teacher, supported by standardised assessment tasks and group moderation.

Aggregation for different reporting purposes

81. Confidence in reported results increases as the number of measures upon which they are based increases. Each of the purposes of the national assessment system can be met by a separate report based on a different aggregate to which a different degree of confidence can be attached. These aggregates can be at class, school, LEA and national level.
82. Formative assessments would take the form of a pupil profile, with each profile component based on a teacher rating and combined with items of information from the standard assessment tasks. A pupil's profile may change over time but this is not a cause for concern. It is indeed entirely proper: the very intention of formative assessment is to provide the information which teachers need to bring about *change* in the pupil's achievements.
83. The evaluative purpose of the national assessment system would be met by aggregations

over large numbers of pupils. Sufficient confidence would attach to these aggregations, target by target, to provide formative feedback both to inform schools and others about their work and to guide those responsible for developing the national curriculum. When targets are further aggregated into profile components there should be a high degree of confidence in conclusions about the delivery of the national curriculum in most schools.

84. These reports are likely to be too complex to be used in facile comparisons of school performance. If, however, profile components are reported on a common scale, the aggregation over pupils and over profile components will yield data which might be used by parents to evaluate schools. Because of the aggregation there can be confidence in the measurement precision of this report. But the aggregation also means that only a very general interpretation of these results is possible. The measurement precision does however have an advantage: the distribution of achievement in a school should be interpretable with some confidence, broadening the basis of evaluation from the current narrow emphasis on average achievement.

We recommend that wherever schools use national assessment results in reports for evaluative purposes, they should report the distribution of pupil achievements.

85. The issue of reporting results at school level is further discussed in Section XII.

Cost effectiveness

86. Our proposals for the national assessment system have been strongly influenced by the importance of securing confidence in the resulting reports. At the same time they are compatible with considerations of economy.
87. The heaviest reporting load would fall upon class teachers of the 11 year old group, who may have to report on more than a dozen profile components per pupil. Such teachers already carry a heavy responsibility in communicating pupil performance to the receiving secondary school. The more structured approach of the national assessment scheme, with its greater degree of external support, should go some way to offset the increased reporting load, and absorb much of the effort currently being taken up by unsystematic assessment.
88. Much of the assessment activity would develop from existing good practice. Teachers' ratings of profile component performance would be based upon whatever internal evidence is currently used by the school when reporting to parents, though with the greater clarity stemming from defined attainment targets.
89. The externally provided assessment tasks would, at the reporting ages, supplement and in part substitute for classroom tests designed by the teacher. The tasks should help to disseminate a sound methodology for probing pupils' achievements for formative purposes. They should also help to communicate national achievement data for the age group, and a range of differentiated targets above and below these expectations.

90. The sequential assessment strategy described above, coupled with a staged progression of targets, would simplify the design of the tests. Each could be focused on a single stage and probe half a dozen profile components. Designed as normal curriculum material, they could be distributed across a term's work with little additional demand on pupils' time. These externally devised assessment tasks, although expensive to produce initially, would have a utility beyond the reporting ages in a staged system. Teachers could use them when appropriate in intervening years to support their judgements of progress conveyed at meetings with parents. In these ways we envisage that the national assessment system will support teachers in every year of schooling, not just the four reporting years.

THE ASSESSMENT SYSTEM IN PRACTICE

XI. Progression, scales and ages

91. As emphasised in our Introduction (paragraph 5) we regard it as one of our priorities to ensure that criteria and scales used should relate to expected routes of development, giving continuity to each pupil's assessment at different ages, and thereby giving a detailed picture of each pupil's progress.

Ages of assessment and reporting

92. To provide a clear basis for subsequent discussion, we first consider the ages at which national assessment and reporting should take place. For national purposes we believe that the balance of advantage lies with reporting towards the end of the school year in which a cohort of pupils becomes 7, 11, 14 or 16. With very few exceptions, the ages of 7 and 11 come either at the end of or, in middle school systems, at least two years into a phase of schooling. By those ages children should therefore feel settled in that phase, and the teachers should know them well. It is important that 7 year olds who have not made a satisfactory beginning in learning to read, write and calculate should be identified and helped to make progress. Where there is transfer, whether after 7 or after 11, the process of reporting should allow useful information to be passed to the next school. The age of 14 is desirable because it is commonly the time when decisions are being taken about the courses children should subsequently follow; and 16 allows a summative assessment to be made for those children moving out of the school system, whether to work or to further education or training.

We recommend that the ages for national assessment should be 7, 11, 14 and 16, with reporting occurring near the end of the school year in which each cohort reaches the age involved.

Progression

93. Given a framework of reporting at those four ages, we go on to assume progress to be defined in terms of the national curriculum, and the stages of progress to be marked by levels of achievement as derived from that curriculum. It is not necessary to presume that the progression defined indicates some inescapable order in the way children learn, or some sequence of difficulty inherent in the material to be learnt. Both of those factors may apply, but the sequence of learning may also be the result of choices for whatever reason, which those formulating and operating the curriculum may recommend in the light of teaching experience.
94. The consultative document on the national curriculum, published in July 1987, proposed that 'the main purpose of assessment will be to show what a pupil has learnt and mastered...' (paragraph 28); that 'attainment targets will be set' for the core subjects, establishing what children should normally be expected to know, understand and be able to do at around the ages of 7, 11, 14 and 16, so enabling the progress of each child to be measured against established national standards (paragraph 23). The subject working groups should, among other things, take account of the need for continuity and progression

throughout compulsory schooling (Annex A, 3.I.c).

95. The profile components described in section VI, as defined by the subject working groups, will identify the main areas of attainment within which progression is to be assessed. In the following paragraphs we consider how information about the pupil's stage of learning can best be conveyed.

Communicating information about stages of learning

96. Most straightforwardly, one could simply describe what a pupil understands, knows and can do in terms similar to those that will be used by the subject working groups to define attainment targets. We believe that it would be right for much of the communication between teachers and parents to take that form at the reporting ages and between them. However, two considerations make it advantageous also to convert the descriptions into a marking scale. One is to convey some sense of where a child is in the process of learning; the other is to make easier the analysis of results from groups of children. The last is of interest to parents who wish to know where their child stands in relation to others of about the same age and also, at various levels of aggregation, to teachers, heads, LEAs, the Government and the wider community.
97. Commonly used systems of scaling pupils' achievements divide those of about the same age into 5 or 6 groups. In England and Wales, some schools use a five point scale: A to E, for example. In the Federal German Republic a 6 point scale is common, and the numbers 1 to 6 are used. Generally the scaling is based on assessment, including the results of tests, which judge a pupil's performance against that of others in the group, or some comparable group: that is, the judgements are norm-referenced. Furthermore, the reference groups are usually of pupils of the same age group. Commonly, in the five-point system the scaling is so arranged that small proportions of pupils are likely to be in grade A and E, larger proportions in Band D, and perhaps half of the children in C. These proportions may vary when the groups being assessed are untypical of the general population. It is, of course, possible deliberately to arrange for the proportions to be different, and a five point scale could be adjusted, for example, to divide the pupils so that very few are judged to be 1 or 5 and the rest are spread evenly between 2, 3 and 4.
98. Levels of achievement could therefore be translated into a five- or six- point scale at each of the reporting ages. This could have the attraction of apparent uniformity across all ages. However, age-specific grading has drawbacks. Pupils are more likely than not to remain in the same grade or level, plus or minus one, at each reporting age. A child might be assessed to be on level 1 at one reporting age, on level 2 at the next and back to level 1 at the third, despite, in absolute terms, having made progress. Even if the child remains on the same level from age to age, he or she will be given no sense of having made the progress which must in fact have been made. Furthermore, even though level 3 at 7 years of age would indicate that a child had achieved specific competences, those same competences would merit a different level number for a slower child, only 1 or 2, at the next reporting age. There would be no simple relation between the level numbers and performance.
99. More generally, the combination of a norm-referenced system with age-specific scaling

would not be consistent with the proposals in the national curriculum consultative document. The overall national purpose is to work for achievement of the attainment targets of the curriculum. Assessment, whether for feedback to pupils or overall reporting and monitoring, should therefore be related to this attainment i.e. it should be criterion-referenced. Given this, it follows that different pupils may satisfy a given criterion at different ages: to tie the criteria to particular ages only would risk either limiting the very able, or giving the least able no reward, or both.

100. In view of these disadvantages of age-specific scaling, we consider that the scaling system should be directly related to the development of pupils' competences as described by the attainment targets of the national curriculum. That is to say, scales should be used that indicate where a pupil has reached in a profile component. We shall use the word level to define one of a sequence of points on a scale to be used in describing the progress of attainment in the profile component. The sequence of levels represents the stages of progression.
101. A pupil assessed as achieving a given level, say level 2, will have satisfied the criteria for level 2 and will be working towards the criteria for level 3. Progress is marked by achievement of successive levels over time. For the purposes of national assessment, a broad indication of progress is required. We propose that the criteria defining successive levels be so chosen that a pupil could reasonably be expected to progress by one level in two years of work in that profile component. Over the age range 7-16 this would imply a need for 5 or 6 levels. At both ends of the age range, however, some will be unable to progress as fast as others, and some will be able to make quicker progress. This leads us to recommend that a total range of ten levels will need to be defined: fuller justification of this number will be given below.

We recommend that each of the subject working groups define a sequence of levels in each of its profile components, related to broad criteria for progression in that component. For a profile component which applies over the full age range 7 to 16, there should be ten such levels, with corresponding reduction for profile components which will apply over a smaller span of school years.

Progression through the levels

102. The implications of this proposal may be seen from the illustration in Figure 1. The details of the representation in this diagram will be explained in three steps below, considering respectively age 7, age 16 (with GCSE) and then intervening ages.

Age 7

103. The first step in the explanation applies to age 7. The range of levels likely to be achieved by the youngest children will be narrower than the range likely to be achieved by the oldest. For reasons that are outlined in section XIII, we believe this to be advantageous, and we recommend that, for reporting purposes, the levels for the 7 year olds should be confined to levels 1, 2 and 3. These levels should, for this age group of children, be regarded as mainly diagnostic in the sense that level 1 identifies those children who may need more help in

their learning than can be provided from ordinary classroom resources; those in level 3 have made such advances that they also need additional help if they are to maintain their speedier progress. We recognise that assessment at levels 1 and 3 will be more common in some aspects of learning than others; example, there are likely to be more children in level 1 for literacy and for mathematics than for some other aspects of learning.

We recommend that levels 1 to 3 be used for national assessments at age 7.

Age 16

104. To develop the second step of the explanation of Figure 1, we note that at the other end of the age range, there has in the past been a variety of attempts to derive written criteria from GCE and CSE grades. That experience suggests that no more than four sets of levels can be identified in relation to the GCSE population.

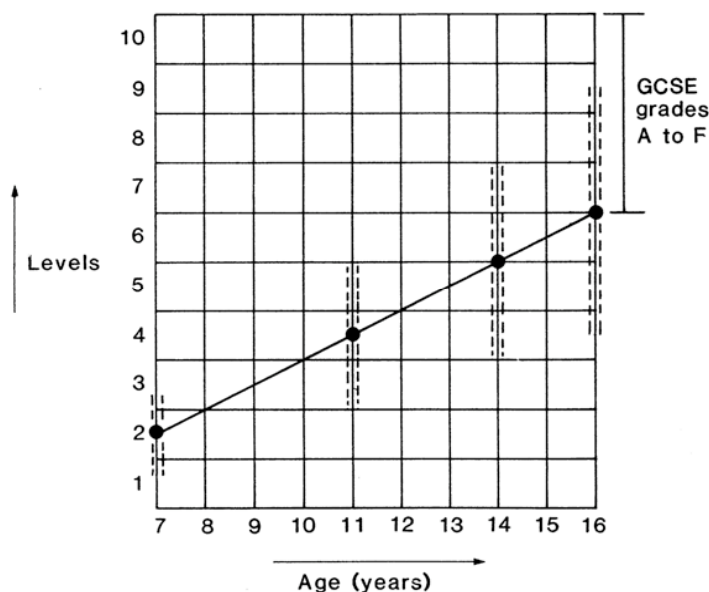


Figure 1: Sequence of pupil achievement of levels between ages 7 and 16

The bold line gives the expected results for pupils at the ages specified. The dotted lines represent a rough speculation about the limits within which about 80% of the pupils may be found to lie.

105. We are conscious of the tendency for new assessment systems to overlay rather than to replace existing schemes, resulting in a growing burden on the schools. It would be unfortunate to have two independent systems in operation for the 16 year old age group. On the other hand, problems of comparability are not readily solved and a premature attempt to link the national assessment system with GCSE could lead to a top-down approach which placed undue constraints on the design of the national curriculum at earlier ages. It is our expectation that the phasing of the introduction of national assessment (see Section XVII) will work towards GCSE rather than away from it.

We recommend that the formal relationship between national assessment and GCSE should be limited, in the first instance, to this one reference point: the boundary between levels 6 and 7 should correspond to the grade F/G boundary for GCSE.

106. The four levels from 7 to 10 could, following this recommendation, bear some relationship to upper GCSE grades, but we believe it would not be helpful at this stage to attempt to specify equivalences between the two systems whilst they operate in different ways. These upper levels should be set by criteria which specify attainment targets appropriate to the GCSE population at age 16, although again this relationship cannot be precise.

We recommend that as they develop the upper four levels of their profile components, the subject working groups adopt present practices for determining GCSE grades at A/B, C/D, mid-E, and F/G as a starting point.

107. It has been suggested by some that GCSE should be revised and extended to provide appropriate targets for the entire age group. We feel this would be unwise on three grounds. GCSE is a major educational innovation which has led to considerable upheaval in the secondary school system: teachers need a period of stability to come to terms with these changes. Secondly, if the national assessment system meets its objectives, a larger proportion of the age group will be served by the existing GCSE system. There is value in setting up a dialogue between the GCSE system and the curricular approaches characteristic of earlier years of education and the time scale allows for this. Finally, in a system of progression, extension of GCSE in the short term is unnecessary, since the levels designed for the middle range of 14 year olds will present appropriate challenges for pupils who at age 16 have not yet progressed into the GCSE grade range.

We recommend that GCSE be retained in its present form until the national assessment system is initiated at earlier ages.

Ages 11 and 14

108. To develop the third step in the explanation, we return to the assumption that a level should be roughly equivalent to two years of educational progress in a profile component. The average expectation for 14 year olds will therefore be the level 5/6 boundary. On similar grounds the average expectation for an age 11 pupil will be level 4. In mathematics, the widely recognised seven year gap in attainment between the most advanced and the slowest pupils at age 11 would imply that levels 3, 4 and 5 should be characteristic of profile components in that subject for the age group but that there will be a few pupils who have already progressed to level 6, the highest level of report for junior school pupils. In general, however, level 6, which should give an indication of how levels might be interpreted in relation to GCSE, should be achieved by some pupils at 14. Level 6 at age 14 is indicative of a pupil who is on target for a GCSE grade. At earlier ages, however, not all of the requisite knowledge and experience may be in place. Furthermore, GCSE grades involve combinations of profile components, and some components may be encountered only late in the curricular experience of the pupil.

109. In reviewing the above argument, it should be emphasised that the basis is the definition, by subject working groups, of a sequence of levels so tied to criteria that two years' learning represents one level of progress. The spread that may be found at any one age, given such a system, remains to be discovered, but the above speculations have some basis in practical experience. For instance, a growing body of evidence indicates that in English and mathematics there is a very broad range of achievement in any year group, sufficiently broad to ensure that there will be overlapping achievements at ages 7 and 11, 11 and 14, 14 and 16. Thus a curriculum target designed for one age group may be quite appropriate for some pupils in an adjacent age group. The need for differentiation at any particular age, coupled with the overlap of achievements between reporting ages, supports our proposal for a single sequence of levels across the age range for national assessments. Indeed in our view the differentiation of attainment targets at a particular age arises naturally from the notion of age progression upon which the reporting scales of national assessment are based. In a criterion-referenced system there should be no distinction between the definition of the sequence of levels of a profile component to reflect progress between ages, and its use to differentiate progression at a particular age. Only one set of criteria is required. The levels defined by the national curriculum attainment targets will provide differentiated challenges at each age according to the needs of the individual pupil. All pupils should then have an expectation of making progress in every profile component: progress which will be indicated by the achievement of new targets.

Differentiation

110. This leads naturally to a discussion of differentiation. As we shall argue in a later section, we think that for the youngest age tests should be devised that, from the children's point of view, are similar in form to their normal school work, should be based on familiar material, and should allow children to exhibit different levels of knowledge, understanding, and skills in their responses, so as to take account of the varying maturity of pupils and the burden of testing on the class teachers, considerable restraint in the number of assessment tasks will also have to be exercised at the 7 and at the 11 year old stage, and this is further discussed in Section XIII.
111. At the later reporting ages it may not always be possible to rely on differentiation by outcome in this way, and the tests devised may themselves have to be sequenced according to the targets being tested, though within the same profile component.
112. Given the view that the differences in performance at any one age must be appraised by the same criteria as progress between ages, it follows that any pupil who has made remarkably fast or remarkably slow progress may well be assessed by the methods used for normal pupils at much older or much younger ages respectively. Thus it would not be impossible for a pupil at age 7 to be capable of level 4 in some components. With this strategy, which is easy to implement because of the cross-age sequence of levels, there would be no need, for the purposes of reporting within the national assessment system, for pupils to take tests or other assessments at earlier or later ages than those specified.

We recommend that assessment and reporting for the national assessment system be at the same ages for all pupils, and that differentiation be based on the use of the single sequence of levels set up to cover progression over the full age range.

Improvement over time

113. The normal distributions across levels at each age will indicate that the attainment targets corresponding to those levels constitute the range of targets appropriate to that age. But as curriculum and teaching are adapted to the new system, improvement in these distributions may be expected: these can be recorded and investigated in our scheme because it is criterion-referenced and not tied to limited expectations at particular ages. Nevertheless, for the reporting and subsequent analysis to be effective, it is important that virtually all pupils be assessed at the same age. It would be possible in any analysis to take account of the fact that (say) 2% of pupils at age 14 achieved level 9 in a particular profile component (and the means for them to exhibit this would be available); it would not be possible to feed into analysis the fact that they were allowed to attempt assessments for level 7 at age 12 and, having succeeded, were excused at age 14.
114. For reporting at a particular age, it will be possible to say of pupils in a particular class, or school, or LEA, what proportions were in each of the levels of the sequence. As recommended above, that set of proportions would constitute the form of report for each profile component.

Special needs

115. In considering the guidance appropriate for individual pupils, the teacher should be giving particular attention to cases of special need. For example, any children who do not move into a higher level between reporting ages should certainly have been identified as needing extra help, as may some who move through the level system slowly. There may be particular difficulties with pupils who might not be able to undertake the normal assessment tasks. All children who are assessed in the agreed profile components should take part in the nationally prescribed tests unless the headteacher certifies that a child with special educational needs is, by reason of those needs, unable to participate in the test. Some children may be able to take tests in an adapted form or with special help – which should be taken into account in determining the results (see Section XV). Children who are statemented will be exempt if their statements exclude their participation in the relevant aspect of the national curriculum. (See Section XV for further discussion of special needs.)

Continuing assessment between reporting ages

116. Whilst most of the above discussion has concentrated on assessment and testing at the four specified ages, it will be necessary to set this programme in the context of the frequent assessment of pupils' work that teachers make, and will continue to make. This work is far more extensive than that of the national assessments, covering progress over all of the periods intervening between 7, 11, 14 and 16. The system of levels tied to criteria which we propose should provide a valuable framework for relating this work to the attainment targets of the national curriculum. A training programme should be arranged to help teachers to become more effective in the frequent assessments that they make, and to help relate these to the criteria embodied in the national tests and assessments. This need was envisaged in paragraph 74 of the consultation document on the national curriculum.

We recommend that support items, procedures and training be provided to help teachers relate their own assessments to the targets and assessment criteria of the national curriculum.

117. Teachers will also need a varied set of diagnostic tests to help explore special difficulties shown up in the national and other assessments.

We recommend that a review be made of the materials available to schools for detailed diagnostic investigation of pupils' learning problems, and that the need for extra help with production or advice about such materials should be considered.

These recommendations should be considered with the complementary recommendation made in paragraph 78.

Assessment of new profile components

118. Most of the foregoing discussion has been set out on the assumption that it is about profile components that are applicable across the full age range from 7 to 16. We recognise, however, that it may be necessary, in some aspects of the curriculum, to introduce new profile components as the children progress, and that these new components may be suitable only for children who have already acquired certain levels of knowledge and understanding. It is for the subject working groups to recommend whether and about what, this should be so. The initial level for these wholly 'new' profile elements should be level 1. A modern foreign language is a special case, since under the national curriculum this would not normally start being taught until the secondary school stage; the reporting system for *all* profile components in a foreign language will need to begin at level 1. We regard this clear divorce of level number from age to be advantageous rather than otherwise, since it will emphasise the criterion-referenced nature of the assessment.

Number of profile components assessed at primary stage

119. Whilst it follows from the above that the number of profile components to be assessed at 7 and 11 will be less than at 14 and 16, there is still a potentially serious problem because of the size of the burden that could too easily be placed on teachers in the primary phase. If the subject working groups were to work independently in making recommendations for primary pupils, then a single class teacher of 7 or 11 year olds could have to assess pupils in at least 20 and possibly 30 or more profile components. Furthermore, some components of foundation subjects would be inappropriate, in their fully differentiated form, at least at age 7.
120. One solution would be to limit assessment in the primary phase to the three core subjects. This could have the effect of narrowing the primary curriculum to an undesirable extent, even if it were only adopted for age 7.
121. There is a different approach to this problem which could be more fruitful. As recommended earlier, subject working groups should pay particular attention to profile

components which might overlap, or be identical, between different subjects because they represent important cross-curricular areas of knowledge understanding and skill. We expect that some profile components will have a wide range of applicability across the subjects of the curriculum, so as to encourage pupils, teachers and parents to recognise common and overlapping aspects in the work being reported. In addition some profile components which should be separately identified in separate subjects in the secondary phase could well be combined up to age 11 because the aspects which differentiate them are not relevant at younger ages. Thus profile components at older ages might evolve by differentiation from a smaller number at younger ages. Figure 2 illustrates this approach.

122. Whilst this approach reduces the numbers of profile components at 7 and 11, it does not neglect the emergence of subject identification at age 11, indeed it allows the context in which assessments and tests are set to become increasingly identified with curriculum subjects as pupils mature. One most important advantage of the scheme represented by Figure 2 is that it stresses the need for continuity in assessment, and therefore in curriculum planning, across the whole age range from 5 to 16. Thus, work at age 11 should both have a planned relationship to the work suitable for 7 years olds and also provide a suitable basis for more highly differentiated work in the next phase. The information provided from national assessment at 11 would then be relevant to the interests of all secondary subject teachers in appraising the starting points for work with pupils after age 11.

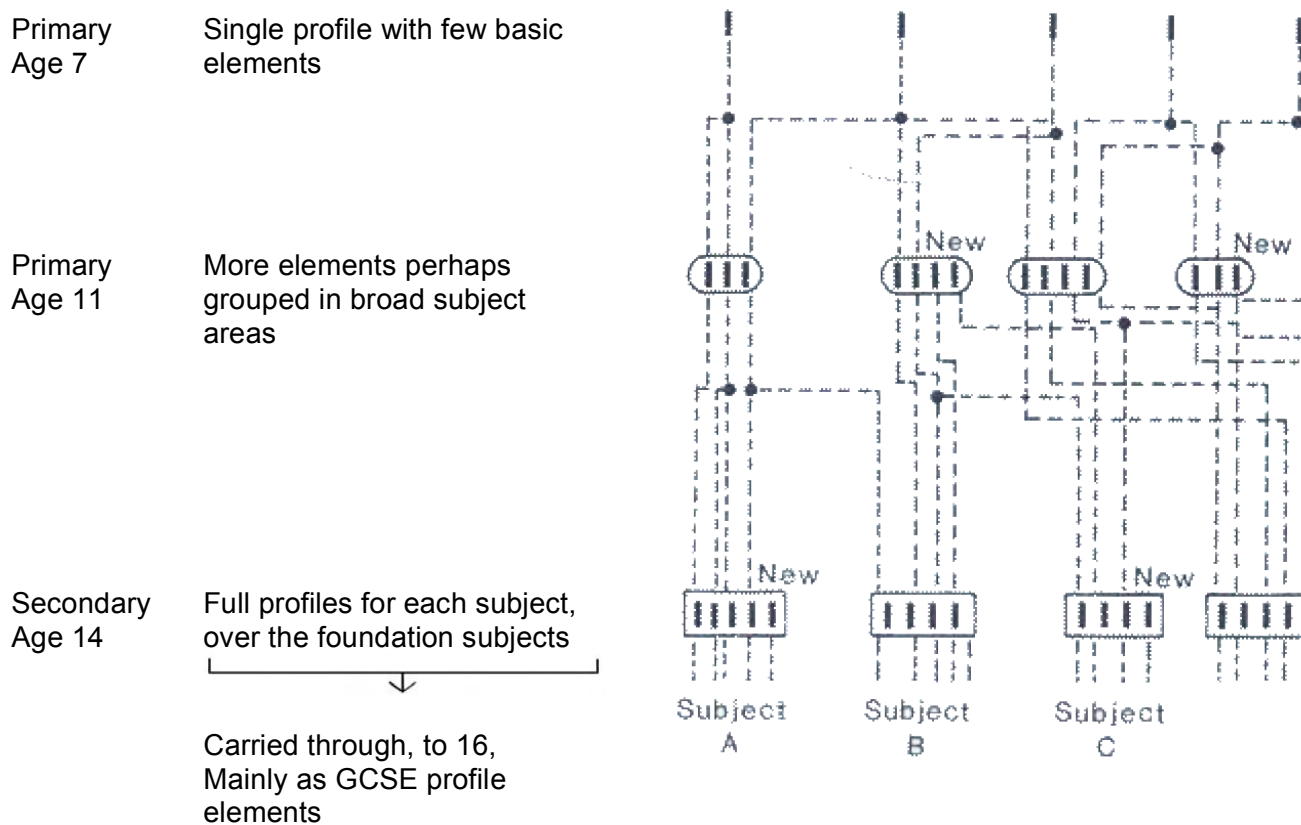


Figure 2: Evolution of profile components

123. It is not possible for us to develop this theme beyond the very general form sketched in Figure 2. The underlying issue is a curricular one and it is for the subject working groups to collaborate with one another in their recommendations for 7 and 11.

We therefore recommend that a working group be established, with some shared membership between the subject working groups, to co-ordinate their proposals for assessment, including testing, at the primary stages, in the light of a comprehensive view of the primary curriculum and of the need to limit the assessment burden on teachers.

124. In summary, progression would be identified through the reporting system on the basis of identifying changes in what the children know. Many of those changes would be denoted by the changes of level within a profile component. Some components would be subdivided as learning progresses, and the level numbering would continue. Some might be wholly new and the level system would commence from level 1. New components and some subdivisions of components may be for use only with children who have reached pre-defined levels or, in the case of a modern foreign language, the appropriate age. At the first reporting age, some profile components would be general and not specific to subjects, but would form the basis for a growing number of subject components which would be developed from them at later ages. Thus progression would be firmly emphasised across the transition from primary to secondary work.

XII. Reporting

125. The results of national assessments will need to be reported to a variety of audiences; the intended audience will determine the degree of detail and interpretation required.

Parents and individual pupils

126. Parents require information about their child's progress in relation to the attainment targets of the national curriculum in a variety of areas and presented in such a way that the individual child's progress can be related to national performance. It is important that results communicated to parents and children should be positive – indicating what the child has achieved; and that they should be constructive – pointing to what needs to be done if the strengths and weaknesses identified by assessment are to be dealt with appropriately. Assessment has to be designed to help all children to proceed with their schooling to the limits of their ability; weaknesses have to be recognised, and appropriate help made available.
127. Thus reports to a pupil and to the pupil's parents should be in terms of profile components since these give the necessary detail about strengths and weaknesses in order to guide future work. As Records of Achievement are more widely implemented, the profile component results should be incorporated into the record, which will also contain other types of information about the pupil (see Appendix I).

Reporting within the school

128. It should be for the teachers within a school to decide how profile component data should be aggregated to provide a basis for internal review. Data aggregated at class level over separate profile components might be of value to a teacher co-ordinating work in an aspect of the curriculum in a primary school, or within a subject department in a secondary school, whilst aggregation over components to give single subject scores might be appropriate in reviews of the school as a whole. Data should show the distributions of performance rather than a single mean and could be compared with the overall national performance.

Employers and further education

129. Results of national assessment of individuals at age 16 may also be taken into account by, for example, employers making decisions about offers of employment, apprenticeships and YTS places and, among others, sixth form and FE colleges making decisions about suitable courses. If the results of assessment at age 16 were included in the pupil's Record of Achievement, employers and colleges would receive a wider view of a young person's potential than could be provided by the results of the national assessment system alone.

Educational decision-makers

130. School governors and local and national providers of education will also need aggregated information about the distributions of performances, within classes and within different subjects. Local and national authorities would have to specify the forms and levels of aggregation that they require. For these audiences scaling or adjustment of distributions should not take place. Parents should also have access to results for the overall performance of the school which their children attend.

Confidentiality for individual pupils

131. If national assessment scores for a school are to be reported to a wider audience than to individual parents, governors and providing authorities, then several problems will have to be considered. It should first be clear that results for individual pupils are confidential, to be shared with each pupil and with his or her parents, and to be transmitted within the school, to a receiving school on transfer, and, if necessary, to the local authority, only in confidence.

We recommend that national assessment results for any individual pupil be confidential, to be discussed between pupil, parents and teachers, and to be otherwise transmitted in confidence. National assessment results for a class as a whole and a school as a whole should be available to the parents of its pupils.

Reporting a school's performance

132. The general problem is that a school's "performance" can only be judged by taking fairly account of many aspects of its work and of many factors outside its control that affect its work. Performance in the national assessment will only be one element in this broad

picture. So any report of assessment results should be part of a broader report covering many activities and achievements of the school, including information that the school may or should publish about its curriculum.

We recommend that the *only* form in which results of national assessment for, and identifying, a given school should be published is as part of a broader report by that school of its work as a whole.

133. It is well known that a variety of socio-economic factors so affect pupils' capacity to respond to school work that they can lead to very wide differences between the performances of different schools (see Appendix J for a review of the evidence). It is difficult to see how such factors can be taken into account in reports of the performance of individual schools. A number of issues arise. Firstly, should the new data be scaled to account for such factors before being presented? Adjusted figures, by scaling up the scores of pupils with low attainments in areas of socio-economic deprivation, could conceal large numbers of pupils who are under-achievers and could also conceal the need for additional resources to be made available to these schools. On the other hand, a misleading view of a school's performance could easily result from the use of unadjusted data as a basis for judgement. For example, average attainments from pupils drawn from a favoured area might appear satisfactory but would in fact indicate that the school was under-performing in relation to the capacities of its pupils. We take the view that the publication of school results to a wider audience than parents, governors and providers and the use of such results to compare schools' performance would be liable to lead to complacency if results were adjusted, and to misinterpretation if they were not.

134. Reports to the public of national assessment results should therefore always be by means of a report produced by the school, authenticated by the local education authority in the case of those schools for which that authority is responsible. Results should not be adjusted for socio-economic background, but the report should include a general statement, prepared by the Local Authority for the area, of the nature and possible size of socio-economic and other influences which could affect schools in the area.

We recommend that any report by a school which includes national assessment results should include a general report for the area, prepared by the local authority, to indicate the nature of socio-economic and other influences which are known to affect schools. This report should give a general indication of the known effects of such influences on performance.

135. We are agreed that the age 7 results of national assessment would not be a useful guide to the quality of education in a school, because of the large and variable effects both of home background factors and of differences in entry ages and pre-entry experience of pupils. This latter factor is discussed in more detail in the next Section. We therefore believe that there should be no requirement on a school for publication of the school's age 7 results to a wider audience than parents, governors and providers. Nevertheless, if in spite of the reasons advanced above, an individual school wished to proceed with the publication, then it should be allowed to do so at its own discretion.

136. A characteristic of primary schools, as distinct from their pupils, that needs to be taken into account in settling on the reporting system for the national curriculum is their size. Most primary schools are small as compared with secondary schools. (See Appendix K on pupil numbers). In particular, the number of pupils in an age group in most primary schools is below 30, and in a significant proportion of schools it is below 10. The number is frequently too small to allow any general conclusions about a school to be drawn from the levels of achievement reached by a single age group, and it may be necessary to consider results over a number of years. That also presents difficulties. Over even as few as 5 years, the staff of a school might have changed considerably and so, probably more often in inner city than in rural areas, may the nature of its catchment area. The aggregated results need to be interpreted with great care.
137. For primary schools with age 11 pupils, we consider that data on national assessment at age 11 should be provided only in the context of a school report as set out above; if schools so desire, data may also be provided on the same cohort at age 7, if these were in the same school at that younger age. These reports would offer general guidance to readers about catchment area effects and be available to any enquirer, although written particularly to guide prospective parents.

We recommend that national assessment results, for pupils at age 11, aggregated at school level, should be published as part of each primary school's report. We recommend that there should be no requirement to publish results for pupils at age 7.

138. For a secondary school, the publication should include its GCSE and other national assessment results for age 16; its national assessment results for age 14 (in both cases showing the distribution of such results); and, if the school so desires, information about progress made by the 16 year old cohort in relation to its national assessment results at age 14. A school with age 11 entry might also, at its discretion, refer to overall data for the distribution of national assessment results for its entry, but this should be done without exposing data on individual feeder schools.

We recommend that national assessment results for pupils at ages 14 and 16, aggregated at school level, should be published as part of each school's report.

XIII. The primary stage

139. We turn now to the implications for work in primary schools of the recommendations we have made.
140. The idea that children's achievements should be described according to what they know and can do is central to the recommendations made in this report. The assessment programme depends upon a clear view of what children should be being taught and should learn: the curriculum. We are not satisfied that a sufficiently clear and coherent view of the primary school curriculum now exists in relation to the national curriculum, and the subject working groups will need to work in close harmony if they are, between them, to define such a curriculum with its associated attainment targets.

Training

141. The sizes of primary schools and the educational immaturity of the pupils have consequences for the ways in which they are organised. Primary school teachers are more often than formerly using their special interests and skills to advise their colleagues and to teach children in addition to those in their own class. However, class teaching remains, and will remain, predominant. From year to year teachers do and should move from teaching one age group to teaching another. Indeed, in a substantial number of classes there are children of more than one age group.

142. At least two things follow. All primary school teachers will need to be trained in assessing at least one of the first two reporting ages, and many may need to be trained for both primary ages; moreover the teachers will need additional training in assessment across the full curricular range of the assessment programme. The arrangements for training primary school teachers with regard to the national curriculum will have to take account both of the large numbers of teachers to be trained, and of the considerable load on each if everyone is to operate without easily accessible support in all areas of the curriculum.

Literacy and numeracy

143. We make the assumption that assessment in primary schools must be concerned, inter alia, with the development of children's powers of literacy and numeracy, the first of which has applications well beyond what might be contained within English, and the second of which should certainly not be the limit of work in mathematics. There are many skills and concepts additional to these that are essential to children's learning both in the primary school and later. It is not for us to categorise them, but we recognise their importance. Some, and perhaps many, are of general application, in the sense that they are necessary in the study of a number of subjects. In expressing such judgements, we recognise that we are making assumptions about an overall view of the primary curriculum that might inform the strategy for assessment and testing.

Organisational factors

144. Most children are taught in school systems that arrange transfer from primary to secondary schools at about 11 years of age. Some children move from infant to junior school at 7 and others attend a full-range primary school from 5, or earlier, until 11. There are also various systems with middle schools. In some, children attend combined first and middle schools until 12; in others they change school at 8 or 9 (or 10 in a handful of cases) and then, respectively, at 12 or 13. In some areas there is transfer at 14. Appendix K on school types and numbers gives detailed statistics.

145. We prefer the reporting ages in the primary stage to be such that the children will have had at least 2 years to settle in their schools and be well known by their teachers. We also regard it as important that sufficient information should be available to a receiving teacher or school so that it is possible for that teacher to provide a programme that takes children forward, rather than to spend time attempting to discover what children know. On balance,

therefore, and bearing in mind the preponderance of systems where the transfer age to secondary schools is 11, we have already proposed that the first two reporting ages for national curriculum purposes should be the school years in which children become 7 and 11 years of age.

146. At 7, a number of factors have to be kept in mind. The children may or may not have had pre-school experience in a mother and toddler club, in a pre-school playgroup, in a social services day centre with child-minders, or in a nursery school or class. Where children are admitted to school at and not before the statutory age, some 7 year olds have had 3 years in school and some only 2. In a growing number of areas, children are admitted before the statutory age: they may be admitted at the beginning of the school year in which they will be 5; or at the beginning of the term in which they will be 5; or at some intermediate time. In areas where social behaviours in the children's homes are different from those they find in school, the children have had less time than at 11 to adjust to school requirements. Those from homes where English is not the first language have had less time than at 11 to acquire fluency in English. The reporting age covers children with an age range of a year: the age difference is proportionately greater at 7 than at 11.

Assessment at age 7

147. We conclude that national assessment at the first stage should be aimed principally at identifying children in need of help which is exceptional for a child of that age, whether because he or she is making so little or so much progress in a profile component. Such help may in practice be provided either by the class teacher or in other ways. There is no suggestion here that all such children are candidates for statementing as defined by the 1981 Education Act, or that special help with regard to one profile component implies that special help is required with respect to other profile components.
148. The results of assessments should be recorded on the pupils' school records and could be of value in reviewing subsequent progress. Care should be taken at all reporting ages, but especially at this age, to avoid giving the impression that the assessment is a prediction of future performance.
149. Because of the immaturity of the 7 year old children and the practices ordinarily found in schools, we consider that externally provided tasks should be no more than three in number for each child. They should seem to the children to be part of ordinary school work, though conducted in a standardised way by the teacher. The tasks should not be differentiated according to levels of difficulty, but should be so constructed that children's performances on them can be related to the national curriculum attainment targets. The topics should be selected by the teacher from item banks so that the matter on which the tasks are based will be within the children's general experience. The three tasks should be drawn from an item bank so as to allow the children to demonstrate clearly the stages reached in literacy and numeracy and the range of other skills deemed appropriate within the national curriculum. Appendix D gives examples of such tasks: each illustrates how in the course of work that should engage pupils' interests, opportunities arise for assessing the competence of pupils in sharply defined profile components. Across a balanced set of three such tasks it will be possible to appraise all of the appropriate components on several occasions and in several different contexts.

150. It has to be remembered that a proportion of children, not least some with special educational needs, are not literate at 7. This should be allowed for in the standard procedures laid down for the assessment tasks. Children should not feel a sense of failure because they are unable to respond to an assessment/task, and it may be necessary to allow children who begin to falter to be helped under prescribed circumstances which, clearly, would lead to it being recorded that the help had been given. The requirements of moderation [see Section X] may well make it necessary that schools in a moderating group should agree on common use of one of the three assessment tasks to be used. We suggest that, during the early years of the national curriculum, common choice should be a requirement, to be abandoned if and when it is found that moderation can be effective even though different schools within a group choose no assessment tasks in common.

We recommend that at age 7 the standard assessment tasks for the national assessment should comprise a choice of three prescribed tasks for each child; each task should be designed to give opportunities for systematic assessment of competence in the range of the profile components appropriate to age 7.

151. As at other reporting stages, the results of the tests and the recorded continuous assessments made by the teacher should be considered together when deciding what level a child has reached.

Assessment at age 11

152. By the second age of reporting, that is at 11, the children are likely to be working in a mixture of topic and subject modes. The subject working parties should consider what applies in their field and recommend in what form profile components are presented. We anticipate that there will be requirements in a subject context, and it is important to bear in mind that secondary schools, though they may undertake some aspects of their early work in a thematic form, may want information that can be linked to a subject analysis of the curriculum. We have already emphasised the need to strengthen collaboration between subject working groups to enable this approach to be made effective.
153. There could still be a particular burden upon the teachers of 11 year olds. Even if the assessment tasks were limited to three as for the 7 year olds, they would have to be more complex to permit the children to show the variety and depth of their competencies. The volume of work they would be likely to produce would be substantial and a thorough, standardised assessment of all its components would take a long time, especially when moderating procedures are allowed for. Without more advice from the subject working groups, we can only draw attention to the issue and tentatively suggest ways forward. It might be possible, for example, to concentrate the process of moderation each year on a sub-set of profile components across the assessment tasks or to limit moderation to two of the tasks. It might be possible to include some work within focused tasks or tests that would be relatively highly standardised and thus simple to mark and requiring no moderation after marking. Perhaps the three tasks should focus respectively on mathematical and scientific learning; on literacy and the humanities, and on aesthetics. It may be that the number of tasks has to be increased to four. But we must stress that these groupings are primarily matters for the subject working groups; our examples are directed

only at illustrating the nature of the issues.

We recommend that at age 11, the tests for national assessment should include three or four standard tasks which cover a range of profile components, possibly supplemented by more narrowly focused tests for particular components.

154. Between each of the reporting stages, continuous assessment is a normal feature of classroom practices and this may be supplemented by diagnostic tests. As we have already noted there is a need for a wider range and variety of means of assessment to be developed for use in busy classrooms at the discretion of the teacher.

Summary of primary school procedures

155. In summary, national assessment in primary schools might work as follows (a fuller description is offered in Appendix L).

156. During the year in which the pupils become 7, teachers would be asked:

- on the basis of continuous assessment allied to programmes of study and attainment targets, to arrive at judgements of their pupils across a number of profile components and using three levels of performance;
- to select from national item banks three standard tasks from within the range of work covered, at least one of which would be chosen in common with other local schools;
- using the standard procedures and criteria for administration and for recording of performance, to incorporate the test tasks into their teaching programme so that each pupil is assessed, over three tasks, in the profile components;
- to attend a local group moderation meeting at which they consider together the results of the tests and their assessments, decide the overall distribution, resolve discrepancies and adjust individual levels in the light of the agreed distribution;
- to compile a confidential report for consideration by head and governors on the overall performance set against local and national distributions;
- to determine the levels for each pupil in the light of the profile components and of general information on standards as a whole, devised from the moderation meetings, and to report to parents;
- to enter results in the school records.

157. For pupils aged 11, the same procedures would apply, except that:

- there are likely to be three or four standardised tasks and these may be supplemented by other forms of test, externally provided, for particular profile components;

- the profile components would be more numerous and more complex;
- the information about overall school performance would be incorporated in a published report, related to a statement of the school's context;
- information on individual pupils would be made available in confidence to the receiving secondary school.

XIV. The secondary stage

Assessment at ages 11-14

158. Assessment in the secondary phase must take account of emerging subject emphases at 11, and reflect a very largely subject-based organisation in the pre-14 stage and later up to 16 when it must articulate with GCSE and Records of Achievement (RoA). Profile components will need to draw on a wide and complex range of subject-based knowledge, skills and understanding, though it is desirable to include some general competences in profile components that all subject working groups agree to use in common.
159. While the national curriculum will render decisions about option choice at 14 less crucial, that age still represents an important staging-point because GCSE and RoA offer at 16 a mainly summative form of assessment – given added force, in the case of GCSE, by being a publicly recognised qualification (for many at the end of formal education).

Assessment at age 16

160. At 16, as we have already indicated, assessment related to the national curriculum will need to articulate with GCSE and other qualifications at equivalent level, as well as with RoA. These are an integral part of the national curriculum provisions. Both in their intentions and in form, they underpin a cumulative picture of achievement built up (with interim reporting) on the basis of assessment confirmed and refined over time, and of work done in a variety of contexts.
161. Together the three initiatives offer, we believe, the prospect of more confident and wide-ranging assessment at 16. To the extent that they have a common purpose and focus, they offer the opportunity for a simpler, coherent and complementary approach to assessing and recording progress having a number of advantages: obviating the current duplication of time and work spent on separate aspects of assessment; keeping assessing, recording and reporting within manageable bounds; alleviating demands on teachers; and providing a focus for teacher development, strengthening their role as assessors and relating curriculum and assessment.

Records of achievement

162. In particular the moderation and national testing required by the national curriculum could

furnish a basis for accreditation and validation for some components of the records of achievement. These in turn provide a means of recording achievement through the secondary years (indeed starting at the end of the primary phase), including that related to personal and social development; and all the more valuable if pupils are drawn formally into discussion with teachers and parents about their progress.

We therefore recommend the use of Records of Achievement as a vehicle for recording progress and achievement within the national assessment system.

An articulated system

163. The technical problem of the articulation of the national assessment levels and the GCSE grades might be solved by adopting numbered national assessment levels (1 to 10 as proposed earlier in this report) across the 7 to 14 age range, and retaining the lettered GCSE grading (A to G). For subjects not being taken at GCSE, the national assessment levels 7, 8, 9 and 10 would be made available within profile components, and the other, lower levels used at age 14 might also be needed. However, the rules of combination of levels in profile components into GCSE subject grades are not our direct concern at this stage and we do not presume to specify how the link between national assessment profile components and GCSE grades is to be made. It will be sufficient to specify that at earlier reporting ages levels 6 and 7 signify that pupils are well on course for GCSE grades at age 16. It is not to be expected that attainment targets as elaborated in relation to the national curriculum, and the national/grade criteria attached to GCSE, will "meet" at all points.

We take it as axiomatic that eventually changes will be necessary to the GCSE and other criteria and that changes derived from the development of the national curriculum should have priority in an orderly process of amendment.

Summary of secondary school procedures

164. In summary, national assessment in secondary schools might work as follows (a fuller description is offered in Appendix M).
165. All secondary schools would receive on transfer detailed information on their pupils on the basis of assessment at age 11 against known attainment targets and programmes of study.
166. During the year in which the pupils become 14, teachers would be asked:
- on the basis of continuous assessment of work done (against programmes of study and attainment targets) and of inter-class moderation, to arrive at judgements of their pupils across a number of subject profile components;
 - to administer nationally prescribed tests in the foundation subjects, to mark and then to analyse the results according to prescribed procedures;
 - to attend a local group moderation meeting at which the teachers' judgements and the test results would be considered together and, with the help of samples and information on local and national distributions, to arrive at an overall assessment, and to resolve

discrepancies;

- to compile an overall report related to 11 year old results for consideration by heads and governors for incorporation in a published report on the school related to its context;
- to adjust any individual pupil's result in the light of the moderation exercise, and then report on each pupil's performance for the pupil and parents concerned;
- to record results in a pupil's record of achievement.

167. **For pupils aged 16**, the same procedures would apply but only in respect of the subjects not being taken for GCSE.

XV. The assessment of children with special educational needs

168. Children with special educational needs make up a very diverse group. Indeed, it could be said that all children, at some time in their educational career, exhibit special teaching needs which have to be met. Children's learning difficulties may stem from physical and sensory disabilities, behavioural and medical problems, reading and communication difficulties or different degrees of mental handicap. They may be educated in special schools, special classes or units in ordinary schools, or may be integrated into ordinary classes. They may or may not be statemented: policies with regard to statementing differ from one LEA to another. Some children with physical disabilities may nevertheless be scholastically extremely able and it is likely that, in their case, parents would rightly press for their inclusion in the national assessment programme.

Need for flexibility

169. Given the extremely diverse nature of special educational needs, it does not seem to us that it would be useful to issue blanket statements about the inclusion or otherwise of such children in the national assessment programme, but rather that the programme should have built into it sufficient flexibility to cope with a wide diversity of needs.

Like all children, those with special educational needs require attainable targets to encourage their development and promote their self-esteem. We therefore recommend that, wherever children with special educational needs are capable of undertaking the national tests, they should be encouraged to do so.

We further recommend that a special unit within the chosen test development agency should be dedicated to producing test materials and devising testing and assessment procedures sufficiently wide-ranging and sensitive to respond to the needs of these children.

Diagnostic materials

170. As for the school population as a whole, but even more crucially in this instance, we see the need for the provision of a wide range of diagnostic tests, to be made available to schools for use on an individual basis. These tests would need to be appropriate to children of a variety of ages, to be administered when they reached a particular level of development and attainment. Teachers of children with special educational needs should be able to draw on a wide range of diagnostic materials at any stage in the children's development, to help to determine the next steps to be taken to guide their personal and educational progress.

Coverage of ability range

171. The national tests themselves should be designed so as to be appropriate to children across the whole ability range, modified as necessary for children with particular sensory problems (i.e. blind and partially-sighted children, deaf and partially-hearing children) and those with communication difficulties. Strategies for ensuring that these requirements were met would need to be devised by the test development agencies during the pilot-testing phase of the test development process, but might include, for example, locator test items to direct children to suitable levels of difficulty, or the provision of items which would allow for a differentiated response. The recent development of computerised adaptive tests might well have an application here.

Exemption

172. Exemption from the national testing programme should, we think, be at the discretion of the head teacher if, in his or her professional judgement, testing would be likely to produce such stress that a medical or emotional condition would be exacerbated. The number of children so exempted should be a matter of record, to be justified during the moderation process. Children exempted from taking the national tests would provisionally be reported as so exempted. Some of them may be identified as in need of further specific help.
173. Some children with special educational needs might be entered for the national tests but might not prove capable of coping with them adequately. To avoid any possibility of such children perceiving themselves as 'failures', the national tests should be designed in such a way as to permit the teacher to curtail the test discreetly without the child being aware that this was being done, or to give assistance, the extent of this being recorded. (Experience with the Assessment of Performance Unit's Practical Mathematics tests has shown that this can be done sensitively and unobtrusively.)

IMPLEMENTATION

XVI. The brief for the subject groups

174. This section draws together the proposals made in preceding sections, provides a review and summary of them, and spells out their immediate implications, in relation to the work of the national curriculum subject working groups.

Framework and exceptions

175. We propose that all the working groups set up for foundation subjects should be expected to operate within the common framework outlined in this report, but that exceptions from one or more of its requirements be allowed if substantial reasons for the exception can be provided. Such exceptions might be permitted more readily for non-core than for core subjects.

Development of profile components

176. Each subject working group should start from its list of attainment targets and combine them to form profile components as described in Section VI above. For level 2, some of the profile components will be broad, becoming differentiated as there is movement through the levels to level 10. Some early profile components might feed into a number of later subject components. In the later levels, the components should relate to such profiles as might be formulated in the development of GCSE grade criteria.
177. Attainment for each profile component should be described in terms of a set of staged criteria, in relationship to which the diverse performances of children can be classified. The levels and corresponding criteria should form a single spectrum irrespective of age, but be so constructed that normal performance of pupils would correspond to the second level at age 7 (for profile components appropriate at that age) and would involve the use of a larger number up to 10 at later ages. These would span the complete range of GCSE grades, as specified in Section XI.

Relations with other subjects

178. It is possible, indeed desirable, that one or more of a subject's profile components might overlap strongly with, or even be identical with components from another subject. Profile components from other subject groups should be studied with this in mind. Cross-curricular themes which are important in teaching should be reflected in overlap, or preferably identity, of appropriate profile components. Special attention should be paid to the need to limit the number of profile components at ages 7 and 11 by use of common components across subjects.

Each subject working group should decide on a limited number, usually 4, of profile components in relation to which any pupil's performance will be assessed and discussed. A criterion-referenced set of levels should be set out for each component, to span the full range of performance over the ages for

which the component is applicable.

Assessment tasks and marking regimes

179. For each profile component, some tests or standard assessment tasks will have to be used, and the brief that a subject working group would want given to any agency responsible for construction and trial of these should be described, including any aspects of the marking which would be particularly significant.
180. Teachers' assessments will also form part of the assessment system. Only broad advice about these is needed, to guide teachers, although details on aspects likely to be unfamiliar might be given so that a case for particular assistance at national level might be considered.
181. For each profile component the methods for combination of standard and internal assessments should be as set out in Section X. If there are any particular problems with the recommended strategy for differentiation – by outcome at younger ages, by task where necessary at older ages – these should be explained.

Moderation

182. We have recommended that the results for each profile component be moderated by the procedures discussed in Section X. If it is felt that group moderation is really essential for some components, but dispensable for others, this should be explained.
183. Recommendations for age 16 will need to relate to the criteria in the appropriate GCSE subjects.

Subject working groups should specify, in broad terms and for each profile component, the appropriate tests (standardised assessment tasks) which should be prepared, and the advice and help which should be given to teachers about their corresponding internal assessments.

Aggregation for reporting

184. Whilst the levels within any one profile component are to be criterion-referenced, it will also be necessary, for reporting to some audiences, to aggregate across profile components to obtain a description of performance in the subject as a whole. The rules for such aggregation will have to be set out: how it may be carried out, whether by simple addition, or by weighted addition, perhaps involving some requirement that minimum levels of achievement must be attained in certain components or by rules which specify the single overall level secured by every one of the possible combinations of profile component levels.
185. The criteria for levels of the separate profile components may have to be related across components if such aggregations are to be sensible. This could be done by reference to national levels of achievement across profile components.

186. Care will have to be taken in determining ways in which results are reported to various users. Some attention should be given to the ways in which the various profile components and the levels within them can be clearly described to the various audiences - pupils, parents, employers, governors and local and national government.

Combination of profile component levels to give a subject level should be by a specified procedure. Uniform ways of describing profile components and the level within each should be specified in language that is helpful to pupils, teachers, parents, employers and other users.

Relations with other agencies

187. It is not possible for us to make precise the boundary between the remit of subject working groups and that of those agencies, including particularly the new National Curriculum and the School Examination and Assessment Councils, who will have to commission construction of standard assessments and oversee operation of the new system. In particular, we note that the specification of resource implications following from the assessment recommendations of subject working groups is not explicitly in their remit.

Novelty and familiarity

188. Nevertheless, the subject working groups will be well equipped to give general advice on two aspects of need. The first is about the degree of novelty of the assessment tasks that need to be constructed: they will know whether what they are recommending for use are well-known and long-established methods, or innovations which will need extensive preliminary trial. The second is about the degree of familiarity, and therefore of confidence, and capability, that teachers may be expected to enjoy in relation to the methods and procedures proposed. Again, general guidance here will be of help in the planning for INSET work that will have to be undertaken.

Subject working groups should give general advice about the degree of novelty of the assessments they envisage, so that the construction of them and the provision of in-service support for teachers can be appraised.

XVII. Phasing

189. It is not possible on assessment grounds alone to set out a firm timetable for putting into operation the scheme that we propose. The assessment and testing is to be constructed in relation to the attainment targets and programmes of study in each subject, and the first determinant of the timetable will be the time required for these targets and programmes to be decided, and for schools to adapt to them.

Secondary schools

190. The situation in the secondary phase is more straightforward than in the primary because it is possible to talk of subjects separately. The following sequences may be envisaged for the

work at age 14 and for work for pupils not taking GCSE at 16 (who will generally use the upper level materials for age 14):

Time	Activity in secondary range
Start of Year 1	Subject Targets Promulgated.
During Year 1	Curriculum. Dissemination. Assessment Materials (at national level) in preparation for age 14.
Start of Year 2	Curriculum set in operation ages 11/12 and 12/13.
During Year 2	Further Development and Trials at age 14. Assessment.
Start of Year 3	Curriculum continues in operation 12/13, 13/14 and starts with new 11/12 group.
During Year 3	Teachers trained, and use assessment methods on trial basis. Some materials trialled for use with non GCSE pupils at age 16.
Start of Year 4	Curriculum in operation years 1-4 (Age 14/15).
During Year 4	First operational use of age 14 assessment with reporting.
Start of Year 5	Curriculum in operation years 1-5.
During Year 5	First operational use of the same materials at age 16 for non-GCSE pupils.

191. This is not a generous timetable. Given the unfamiliarity of the procedures and the need to adapt teaching methods to incorporate them, a trial run year is strongly recommended, at least at age 14. The timetable means that those following the system in the first year of formal operation at age 14 will have been taught within the national curriculum since age 11, although without the primary school background that the national curriculum will envisage. Those first assessed at 16 will have been taught within the national curriculum since age 12 and tested on a trial basis at age 14.

192. It will be important to stagger the implementation of the age 14 and of the age 16 work because the burden will usually fall on the same teachers. If 14 comes first, it becomes more easy to ensure that the first cohort to be involved at age 16 will have followed the national curriculum for 4 years.

GCSE

193. The above takes no account of GCSE developments. For at least the first four years, existing GCSE syllabuses and assessment methods should remain in operation. During that time, any changes in GCSE which follow from the definition of attainment targets up to age 16 in the national curriculum will have to be planned. An outline timetable might be:

- **Year 1:** Agreed subject targets ready for consideration. New GCSE criteria prepared.
- **Year 2:** Examining groups prepare new syllabuses and specimen examinations.

- **Year 3:** New syllabuses and examinations approved and published.
- **Year 4:** Schools start teaching to new syllabus for Year 4 GCSE pupils.
- **Year 5:** First entry to new GCSE examinations.

194. Ideally revisions of GCSE assessment methods should arise from the evolution of the national system. However, a radical change in GCSE seems hardly feasible on the above timetable, which is only just possible on the assumption that the attainment targets and programmes of study for age 16, promulgated after consultation on the recommendations of the subject working groups, mean that the processes of consultation to settle new GCSE criteria can take a much shorter time than in the past. Even given this, the timetable would allow very little time for consultation, or for referral back of particular proposals of the group syllabuses stage. It may therefore be necessary to revise GCSE in two stages, one to incorporate the national curriculum targets within present systems of assessment and grading, followed later by more radical changes to the latter when the rest of the national assessment system had been in full operation for at least a short time. All of the above could be considered as timetables for an individual subject. If the year of commencement (year 1 in the tables) is not the same for all subjects, as seems almost certain, the adaptation in schools will be spread over a longer time.

Primary schools

195. For the primary phase, there is an additional complication that the separate subject proposals ought to be combined, so that primary teachers are working to an appropriate set of profile components as envisaged in Section XI. It seems inevitable that an interim set of profile components should be recommended for 7 and 11 when outcomes from enough subjects are available. Initial curriculum and assessment work would start with these and be adapted subsequently. It should be stressed that in the primary phase, in contrast to the secondary, the addition of guidelines deriving from further subject working groups may imply extra adaptation for the teacher who has already been operating with the outcomes of earlier working groups, rather than mere addition of extra profile components from each group. This could lead to an intolerable load on the primary teacher.

196. Given the above reservation, which applies to the fixing of year 1, the following timetable should be feasible:

Time	Activity in primary range
Start of Year 1	Interim profile component. Targets promulgated for age 7.
During Year 1	Curriculum. Dissemination. National assessment tasks for age 7 prepared.
Start of Year 2	Curriculum in operation ages 5/6. Interim targets for age 11 promulgated.
During Year 2	Trials of assessment tasks, Training starts age 7: tasks for age 11 prepared.

Start of Year 3	Curriculum in operation for ages 6/7. Also started for ages 8/9.
During Year 3	Teachers trained and use age 7 assessments on trial basis. Trials of age 11 materials.
Start of Year 4	Curriculum for all up to the 9/10 age group.
During Year 4	First operational use of age 7 assessment. Trials of age 11 materials and teachers trained.
Start of Year 5	Curriculum for all up to year 5 (10/11).
During Year 5	First operational use of age 11 assessment.

197. This has features similar to those for the secondary phasing. The staggering of ages 7 and 11 spreads out each school's adaptation load and leaves more time for anyone group of children to be taught to the new targets over several years before the assessment system affects them. It also makes it possible to delay to year 2 the decisions about age 11 profile components, so that collaboration between more of the subject working groups can be arranged.
198. The above proposition accepts that pupils will reach the assessment ages with fewer years of experience of work adapted to the national curriculum attainment targets than they should have. For example, all of the first 5 years of secondary work and all of the first years of operation of age 14 and 16 assessments will be on those unaffected by the primary changes. Ideally, a new scheme would be implemented with an age 5 entry cohort and take 11 years to implement as they advance to age 16, but we recognise that the delay this would involve might be unacceptable for achieving national improvements in the work of secondary schools.
199. The times we have given are minimum times. If the implementation were to be more rapid, the work would start with assessment materials of mediocre quality, with pupils who have not had adequate time with the new curriculum, and with teachers inadequately prepared and feeling de-skilled by having to use procedures which they have not been able to assimilate into their own teaching. All of these features would make for a disastrous start and would set precedents from which the new system would take many years to recover. The phasing should pay regard to the undue stress on teachers, on their schools and the consequent harm to pupils which will arise if too many novel requirements are initiated over a short period of time.

We recommend that the new assessment system should be phased in over a period adequate for the preparation and trial of new assessment methods, for teacher preparation, and for pupils to benefit from extensive experience of the new curriculum. We estimate that this period needs to be at least 5 years from the promulgation of the relevant attainment targets.

A SYSTEM OF SUPPORT

XVIII. The support structure

200. The system of national assessment proposed in this report will call for the involvement of a number of agencies.

SEAC and NCC

201. At the heart of the system will be the new School Examinations and Assessment Council and the National Curriculum Council, to co-ordinate the development of the national curriculum and the programme of assessment. It will also be essential for these Councils to arrange to provide advice on assessment to the various subject working groups, to ensure they furnish sufficiently clear statements of attainment targets and profile components to allow appropriate means of assessment to be developed. There would also need to be close liaison between the subject working groups and the chosen test development agencies, probably mediated through the two Councils.

Test development

202. The construction of the national tests would need to be contracted out to appropriate test development agencies. These agencies would have two main functions. The first would be the construction and trial of the national tests and assessment tasks to be used by all. The second would be the development of a range of supplementary tests and assessment procedures to support the educational process generally and to be used at the discretion of schools. The diagnostic tests would be closely focused, applicable at a variety of ages, and would concentrate on analysis of methods and strategies used by pupils in order to identify their problems precisely and so to indicate how they should be helped. For both of these two functions, a range of agencies might be invited to take part. It is possible that some of the development for the second function could be undertaken by commercial test publishers as part of their normal programme.

Delivery

203. The delivery of national assessment and testing items and arrangements, the organisation of moderation procedures, and the collection and collation of results form a complex and interlinked set of functions. Maintenance of the resources, particularly of the collections of tests, whether broad standard assessment tasks or more sharply focussed types of test, will probably have to be a central function. Distribution would include contacting schools, with an up-to-date register of all of them, and a service to deal with enquiries and to help with advice on training in marking and scaling procedures, and help in data analysis. This could be done either nationally by a central agency, or regionally through examining groups and/or LEAs. The moderation work would require organisation at a regional or local level and would have to be linked to INSET programmes for training teachers in the new assessment methods so that they could play their central role. The collection and collation of results might seem to be a routine matter of data analysis, but the outcomes of moderation work would be a valuable source of feedback about the quality of the national

assessments, and such information ought to be carefully collected and analysed. Organisations capable of operation at regional and local level would be needed, and the LEAs or the examining groups, or both in collaboration, might be well equipped to undertake these functions.

Publications

204. To support those operating this system at all levels there will be a need for suitable publications, which the new Councils may have to commission and distribute.

Monitoring

205. The operation of the system ought to be carefully monitored. In particular, the effects of the new assessment and testing procedures on classroom teaching, on the overall curriculum and on existing assessment practices on schools, ought to be a subject for careful review, through commissioned research. A strategy for evaluation of the system and for its improvement, using a variety of sources of feedback, will have to be formulated, probably by the School Examinations and Assessment Council.

XIX. INSET and other resource needs

206. The assessment and testing upon which we have to advise is part of a broader programme of educational change. The introduction of a national curriculum with its attendant attainment targets would, even if there were no implications for assessment, require additions to the facilities and resources of schools, the recruitment of teachers in appropriate distributions across schools and across the subject areas of the curriculum, and the pre-service and in-service education and training of these teachers to ensure that they were able to work well within the revised system. Therefore the needs associated with the changes in assessment have to be placed within this broader context.

In-service training needs

207. It will in particular be necessary for government and LEAs to ensure that appropriate in-service opportunities are put in place. Such opportunities will need to be designed so as to ensure that, first, all teachers understand the new assessment arrangements which support the national curriculum; and, second, teachers who are specifically involved in assessment at particular ages and/or in particular subject areas at both primary and secondary level are equipped with the necessary range of skills and techniques.
208. With the introduction of the GCSE, secondary schools and LEAs have had recent experience of the in-service programmes which of necessity accompany examination reform. The new arrangements for assessment cover the age range of 5-16 whereas the GCSE was essentially limited to the 14-16 age group; but in other respects the two developments have much in common. In both cases a revised view of the curriculum is being developed on the basis of good practice in the recent past, and the curriculum is to be

associated with and supported by a change in the system of assessment. The INSET programme for the GCSE has been scrutinised by the DES and HMI, the SEC and the GCSE Examining Groups, and by the local authority associations. There is, therefore, a wide body of experience on which it would be wise to draw in formulating the larger INSET programme which revised roles for groups of teachers at four ages will inevitably require.

Facing new objectives

209. We draw special attention to only a few issues. It seems that in the GCSE more teachers have encountered problems in carrying out their new functions in assessment than in accommodating to a revised syllabus or to a shift in the definition of the field of study. The use in the GCSE of a training programme which started with the national criteria and only in the later stages of training got down to the detail of individual syllabuses and their assessment systems meant that courses had often started before teachers had been able to think about the immediate concrete problems to be faced in their own schools, using a particular examining body. If, as in the GCSE, there is a shift in the balance of assessment objectives so that skills and abilities are given increased importance, this change alone requires more assessment time, more skill in the assessors and, in many cases, a reduction in the pupil/teacher ratio.

Materials needed

210. From these and similar points, we conclude that it would be unwise to start an INSET programme relevant to a particular assessment system until a range of assessment materials, of all types for the specified age group, could be brought into use in the training course. The types of assessment discussed in Section VII should all be brought under scrutiny and used in exercises in the course. Some of the practices and procedures used in "agreement trials" within the GCSE system could probably be modified to suit this INSET work.

Scale of operation

211. Apart from modifications which might arise for GCSE, the new work required for age 14 would imply a training operation roughly comparable in scale to the GCSE programme, perhaps greater because it would bear on secondary work up to age 14 which has not hitherto worked to constraints of any external system. For the primary phase, the whole operation would be more extensive than the GCSE programme because all the teachers involved will need to be trained in the assessment of all curricular areas. These needs should be carefully borne in mind in considering the phasing of introduction of any programme.

Other training needs

212. Other training needs will also arise. LEA staff of various kinds will need to be thoroughly conversant with the new system as part of their responsibilities in relation to the national curriculum. The management of the whole programme of assessment and the reporting thereof also points to the need for someone in each school (or within a group of small schools) to have the responsibility and skills for such work. The institutions which provide

initial and in-service teacher training will need to be equipped to help with the new developments.

Other resource needs

213. In addition to in-service needs, the introduction of new arrangements will mean that a good deal of new work has to be undertaken if appropriate instruments for assessment and testing at the four age levels are to be available. Beyond the developmental stage there will be costs for the production, distribution and administration of sets of national test materials and of supplementary diagnostic items. At a later stage there will also be costs associated with whatever "reporting arrangements" are determined. It has not been practicable at this juncture in our work to quantify these.

214. We recognise that our proposals will require a substantial investment of time (particularly teacher time) and other resources. The level of investment will be conditioned by whatever decisions are taken on the final shape of the new system and the rate and phasing of its implementation. Detailed analysis will have to be undertaken; in the meantime we identify in Appendix N data about existing and planned provisions which should contribute to consideration of the implementation of a new system.

XX. The local education authority's role

215. We envisage that local education authorities will have a major part to play in implementing the proposals contained in this report. Essentially they will be concerned with the local delivery of the new arrangements and for meeting within their areas the requirements of the nationally determined framework. They will be called upon to play a significant part in the preparation that will be necessary before a new system can be introduced. They will carry substantial responsibility for ensuring the implementation of the new arrangements at school level. They will be engaged in monitoring and evaluating the operation of the new system and in ensuring compliance with specific obligations (e.g. to establish a mechanism for dealing with complaints; to report on and publish information about the delivery of the national curriculum, including assessments).

216. We see LEAs as having responsibility for preparing and overseeing local strategies for giving advice and help to schools in relation to overall policies, and for training and supporting teachers largely through the establishment of specific INSET arrangements to meet the new needs. LEAs will need to work closely with other agencies involved in the delivery of a new assessment system. Decisions on the allocation and deployment of many of the resources which will be needed if the system is to be implemented in a planned and cost effective manner will rest with them. They may have a particularly important part to play, perhaps in co-operation with examining groups, in the arrangements for moderation using group meetings, as therefore both assessment and INSET functions. They will also be well placed to ensure that information about the new assessment arrangements and their outcomes is made available to the range of interested parties – schools will obviously be directly involved with parents, teachers and governing bodies, but on a wider front the general public and employers will need to be able to look to the LEA for this. In undertaking

this task, one of the particular points to which LEAs will need to give attention is how best to set published information in the context of other relevant information about the service in their areas.

217. We believe that the existing strengths of LEAs will provide a good basis on which to build their contribution to the new system. They have at their disposal considerable experience of and expertise in curriculum matters generally and to some extent also in assessment/testing/examining arrangements. Some LEAs have already shown the value of co-operation between themselves and the examining boards and groups in the development and implementation of new forms of assessment. They have administrative and advisory support services available; they have mechanisms for ensuring the supply of books, materials, accommodation and non-teaching support. These existing services provide a structure which can be used for the new task. They will however need to be enhanced. But the extent to which this is necessary will clearly depend on the degree of sophistication contained in the new arrangements, and cannot be considered in isolation from the general range of demands which the totality of the new national curriculum will undoubtedly place on LEA inspectorate/advisory services and on teachers in particular.

XXI. Future work for the task group

218. The problems and prospects considered in the last few sections of our report will require further and more detailed examination than we have been able to give so far. We propose to continue work on them and to produce a supplementary report. For this reason we make no substantive recommendations on these issues at this stage.

SUMMARY AND CONCLUSIONS

XXII. Conclusions

219. Whilst the system we propose draws on many aspects of good practice that are already established, it is radically new in the articulation and comprehensive deployment of methods based on such experience. We are confident that the system we describe is practicable and can bring benefits to work both within schools and outside them. In particular, we can see how provision of new types of support within a framework of a new set of procedures can replace much of the large volume of testing and assessment at present in use. A co-ordinated system will use resources to better effect and will complement and support the existing assessment work that teachers already carry out. Thus the system should contribute to the raising of educational standards so that the broad educational needs of individuals and the national need to enhance the resources and skills of young people can be met.

Building on good classroom practice

220. As we stressed in our Introduction, the proposed procedures of assessment and testing bear directly upon the classroom practices of teachers. A system which was not closely linked to normal classroom assessments and which did not demand the professional skills and commitment of teachers might be less expensive and simpler to implement, but would be indefensible in that it could set in opposition the processes of learning, teaching and assessment.

Formative assessment to support learning

221. Our terms of reference stress that the assessment to be proposed must be "supportive of learning in schools". We reiterate that the four criteria set out in Section 1 are essential if this support is to be secured and we believe that they necessarily follow from the aims expressed in the consultative document on the national curriculum. The **formative** aspect follows almost by definition. For the system to be formative it must produce a full and well-articulated picture of the individual child's current strengths and future needs. No simple label 1-6 will achieve this function, nor is any entirely external testing system capable of producing the necessary richness of information without placing an insuperable load of formal assessment on the child. The formative aspect calls for profile reporting and the exercise of the professional judgement of teachers.

Raising standards

222. The system is also required to be formative at the national level, to play an active part in raising standards of attainment. **Criterion-referencing** inevitably follows. Norm-referenced approaches conceal changes in national standards. Whatever the average child accomplishes is the norm and if the average child's performance changes the reported norm remains the same figure. Only by criterion-referencing can standards be monitored. Only by criterion-referencing can they be communicated. Formative assessment requires the involvement of the professional judgement of teachers. Criterion-referencing helps to inform these judgements. Group **moderation** will enable the dissemination of a shared

language for discussing attainment at all levels – the central function of assessment. These three features will help to emphasise growth. They result in **progression** – a key element in ensuring that pupils and parents receive focused and evolving guidance throughout their school careers. Consistent and de-motivating confirmation of everything as it was at the previous reporting age can be avoided only if pupils and parents can have clear evidence of progress by use of the single sequence of levels across all ages in the way that we propose.

The unity of our proposals

223. We have considered systems of assessment and testing which are very different from the one that we propose. All alternatives impoverish the relationships between assessment and learning, so that the former harms the latter instead of supporting it. Most of them give no clear information or guidance about pupils' achievement or progress, and they all risk interference with, rather than support of, teachers' work with pupils. Thus we cannot recommend any simpler alternative to our proposals. There is of course room for variation in their implementation: for example, using group moderation procedures for a restricted number of profile components, or not using all such procedures on every annual assessment cycle, or phasing in more slowly to spread the load on teachers and on the support systems. None of these marginal changes would destroy essential features of our system, although they might weaken its impact in the short term. However, any major change that we can envisage would destroy the linked unity of our proposals and lose most of the benefits which they are aimed to secure within and for the national curriculum.

Securing teachers' commitment

224. The underlying unity of the three aspects of education – teaching, learning, and assessment – is fundamental to the strategy which informs our proposals. The strategy will fail if teachers do not come to have confidence in, and commitment to, the new system as a positive part of their teaching. Securing this commitment is the essential pre-condition for the new system to realise the considerable value that it could bring.

225. Amongst the conditions which will have to be met to secure this professional commitment will be the following:

- Clear acceptance that the aim is to support and enhance the professional skills that teachers already deploy to promote learning.
- Clear recognition that the focus of responsibility for operation of a new system lies with teachers within schools.
- Stress on the formative aims and on giving clear guidance about progress to pupils and to their parents.
- Widespread consultation and discussion before proposals are put into effect.
- A realistic time-scale for phasing in a new system.
- Adequate resources, including in-service provision.

- Help with moderation procedures so that the system contributes to communication within schools, between schools, parents and governors, and to the community as a whole about the realisation and evaluation of the aims of schools.
- Sensitive handling of any requirements for outside reporting, recognising that simplistic procedures could mislead parents, damage schools, and impair relations between teachers and their pupils.

226. If there is one main motive to explain our support for the system we propose, it is that we believe that it can provide the essential means for promoting the learning development of children: support for teachers in enhancing the resources and professional skills which they deploy.

XXIII. A list of recommendations

227. The recommendations we have made in this report are listed below.

Purposes and Principles

1. The basis of the national assessment system should be essentially formative, but designed also to indicate where there is need for more detailed diagnostic assessment. At age 16, however, it should incorporate assessment with summative functions. *(Paragraph 27)*
2. All assessment information about an individual should be treated as confidential and thus confined to those who need to know in order to help that pupil. *(Paragraph 28)*
3. For summative and evaluative purposes results should be aggregated across classes or schools so that no individual performances can be separated out. *(Paragraph 29)*
4. Assessment of attitudes should not form a prescribed part of the national assessment system. *(Paragraph 30)*
5. To realise the formative purpose of the national assessment system, pupil results in a subject should be presented as an attainment profile. *(Paragraph 33)*
6. An individual subject should report a small number (preferably no more than four and never more than six) of profile components reflecting the variety of knowledge, skills and understanding to which the subject gives rise. Wherever possible, one or more components should have more general application across the curriculum: for these a single common specification should be adopted in each of the subjects concerned. *(Paragraph 35)*
7. The national system should employ tests for which a wide range of modes of presentation, operation and response should be used so that each may be valid in relation to the attainment targets assessed. These particular tests should be called "standard assessment tasks" and they should be so designed that flexibility of form and use is allowed wherever this can be consistent with national comparability of results. *(Paragraph 50)*

8. Assessment tasks should be reviewed regularly for evidence of bias, particularly in respect of gender and race. *(Paragraph 52)*
9. Attainment targets should be exemplified as far as possible using specimen tasks. Such tasks can then assist in the communication of these targets. *(Paragraph 56)*
10. A mixture of standardised assessment instruments including tests, practical tasks and observations should be used in the national assessment system in order to minimise curriculum distortion. *(Paragraph 59)*
11. Teachers' ratings of pupil performance should be used as a fundamental element of the national assessment system. Just as with the national tests or tasks, teachers' own ratings should be derived from a variety of methods of evoking and assessing pupils' responses. *(Paragraph 60)*
12. When the subject working groups provide guidance on the aggregation of targets into a small number of profile components, they should have regard to the need for each component to lead to a report in which reasonable confidence is possible. *(Paragraph 61)*
13. Teachers' ratings should be moderated in such a way as to convey and to inform national standards. *(Paragraph 62)*
14. The national assessment system should be based on a combination of moderated teachers' ratings and standardised assessment tasks. *(Paragraph 63)*
15. Group moderation should be an integral part of the national assessment system. It should be used to produce the agreed combination of moderated teachers' ratings and the results of the national tests. *(Paragraph 77)*
16. An item bank of further assessment instruments should be available for teachers to use in cases where they need additional evidence about particular pupils. *(Paragraph 78)*
17. The final reports on individual pupils to their parents should be the responsibility of the teacher, supported by standardised assessment tasks and group moderation. *(Paragraph 80)*
18. Wherever schools use national assessment results in reports for evaluative purposes, they should report the distribution of pupil achievements. *(Paragraph 84)*

The Assessment System in Practice

19. The ages for national assessment should be 7, 11, 14 and 16; with reporting occurring near the end of the school year in which each cohort reaches the age involved. *(Paragraph 92)*
20. Each of the subject working groups should define a sequence of levels in each of its profile components, related to broad criteria for progression in that component. For a profile component which applies over the full age-range 7 to 16, there should be ten such levels, with corresponding reduction for profile components which will apply over a smaller span of school years. *(Paragraph 101)*

21. Levels 1 to 3 should be used for national assessments at age 7. (*Paragraph 103*)
22. The formal relationship between national assessment and GCSE should be limited, in the first instance, to this one reference point: and accordingly the boundary between levels 6 and 7 should correspond to the grade F/G boundary for GCSE. (*Paragraph 105*)
23. As they develop the upper four levels of their profile components, the subject working groups should adopt present practices for determining GCSE grades at A/B, C/D, mid-E, and F/G as a starting point. (*Paragraph 106*)
24. GCSE should be retained in its present form until the national assessment system is initiated at earlier ages. (*Paragraph 107*)
25. Assessment and reporting for the national assessment system should be at the same ages for all pupils, and differentiation should be based on the use of the single sequence of levels set up to cover progression over the full age range. (*Paragraph 112*)
26. Support items, procedures and training should be provided to help teachers relate their own assessments to the targets and assessment criteria of the national curriculum. (*Paragraph 116*)
27. A review should be made of the materials available to schools for detailed diagnostic investigation of pupils' learning problems, and that the need for extra help with production or advice about such materials should be considered. (*Paragraph 117*)
28. A working group should be established, with some shared membership between the subject working groups, to co-ordinate their proposals for assessment, including testing, at the primary stages, in the light of a comprehensive view of the primary curriculum and of the need to limit the assessment burden on teachers. (*Paragraph 123*)
29. National assessment results for any individual pupil should be confidential, to be discussed between pupil, parents and teachers, and to be otherwise transmitted in confidence. National assessment results for a class as a whole and a school as a whole should be available to the parents of its pupils. (*Paragraph 131*)
30. The *only* form in which results of national assessment for, and identifying, a given school should be published should be as part of a broader report by that school of its work as a whole. (*Paragraph 132*)
31. Any report by a school which includes national assessment results should include a general report for the area, prepared by the local authority, to indicate the nature of socio-economic and other influences which are known to affect schools. This report should give a general indication of the known effects of such influences on performance. (*Paragraph 134*)
32. National assessment results, for pupils at age 11, aggregated at school level, should be published as part of each primary school's report. There should be no requirement to publish results for pupils at age 7. (*Paragraph 137*)
33. National assessment results for pupils at ages 14 and 16, aggregated at school level, should be published as part of each school's report. (*Paragraph 138*)
34. At age 7 the standard assessment tasks for the national assessment should comprise a

choice of three prescribed tasks for each child; each task should be designed to give opportunities for systematic assessment of competence in the range of profile components appropriate to age 7.

35. At age 11 the tests for national assessment should include three or four standard tasks which cover a range of profile components, possibly supplemented by more narrowly focused tests for particular components. *(Paragraph 153)*
36. Records of Achievement should be used as a vehicle for recording progress and achievement within the national assessment system. *(Paragraph 162)*
37. Eventually changes will be necessary to the GCSE and other criteria. Changes derived from the development of the national curriculum should have priority in an orderly process of amendment. *(Paragraph 163)*
38. Like all children, those with special educational needs require attainable targets to encourage their development and promote their self-esteem. Wherever children with special educational needs are capable of undertaking the national tests, they should be encouraged to do so. *(Paragraph 169)*
39. A special unit within a chosen test development agency should be dedicated to producing test materials and devising testing and assessment procedures sufficiently wide-ranging and sensitive to respond to the needs of these children. *(Paragraph 169)*

Implementation

40. Each subject working group should decide on a limited number, usually 4, of profile components in relation to which any pupil's performance will be assessed and discussed. A criterion-referenced set of levels should be set out for each component, to span the full range of performance over the ages for which the component is applicable. *(Paragraph 178)*
41. Subject working groups should specify, in broad terms and for each profile component, the appropriate tests (standardised assessment tasks) which should be prepared, and the advice and help which should be given to teachers about their corresponding internal assessments. *(Paragraph 183)*
42. Combination of profile component levels to give a subject level should be by a specified procedure. Uniform ways of describing profile components and the level within each should be specified in language that is helpful to pupils, teachers, parents, employers and other users. *(Paragraph 186)*
43. Subject working groups should give general advice about the degree of novelty of the assessments they envisage, so that the construction of them and the provision of in-service support for teachers can be appraised. *(Paragraph 188)*
44. The new assessment system should be phased in over a period adequate for the preparation and trial of new assessment methods, for teacher preparation, and for pupils to benefit from extensive experience of the new curriculum. This period needs to be at least 5 years from the promulgation of the relevant attainment targets. *(Paragraph 199)*