APPENDICES

A. Terms of reference and membership
B. Letter of guidance from the Secretary of State
C. Meetings with invited groups
D. Test tasks for 7 and 11 year olds
E. Test and assessment methods: examples of the variety available
F. National testing and equal opportunities: a paper by the Equal Opportunities Commission
G. Technical note on reliability, validity and error
H. The conduct of group moderation meetings
I. Records of Achievement
J. Technical note on background factors affecting educational performance
K. Numbers in schools
L. How the system might work in a primary school
M. How the system might work in a secondary school
N. Resource needs: some data on existing and planned provision
APPENDIX A

Task Group on Assessment and Testing: Terms of reference

1. To advise the Secretary of State on the practical considerations which should govern all assessment including testing of attainment at age (approximately) 7, 11, 14 and 16, within a national curriculum; including the marking scale or scales and kinds of assessment including testing to be used, the need to differentiate so that the assessment can promote learning across a range of abilities, the relative roles of informative and of diagnostic assessment, the uses to which the results of assessment should be put, the moderation requirements needed to secure credibility for assessments, and the publication and other services needed to support the system – with a view to securing assessment and testing arrangements which are simple to administer, understandable by all in and outside the education service, cost-effective, and supportive of learning in schools.

2. In making recommendations, to take into account the need not to increase calls on teachers’ and pupils' time for activities which do not directly promote learning, and to limit costs.

3. To advise on the possible staging of the introduction of assessment including testing to reflect the need for the process to be manageable and for teachers to be adequately trained.

4. To report to the Secretary of State by Christmas 1987.

Task Group on Assessment and Testing: Membership

Chairman

Professor P J Black OBE, KSG, BSc, PhD FIP
Professor of Science Education at the University of London and Head of the Centre for Educational Studies at King’s College, London (KQC)

Members

Professor J T Allanson
Emeritus Professor of Electronic and Electrical Engineering and member of the Secondary Examinations Council

Mr J A Barnes
Director General of the City and Guilds of London Institute

Dr C Burstall
Director of the National Foundation for Educational Research
Mr O Hughes  
Headteacher of a secondary school in Clwyd

Mrs D Kavanagh  
Headteacher of a primary school in Oxfordshire

Mr J O Morris  
County Education Officer, Essex County Council

Mr D A T N Payne  
Director of Personnel, Rolls Royce PLC

Ms H Steedman  
Researcher at the National Institute for Economic and Social Research, London

Mr N Thomas  
Visiting Professor of Education, Warwick University; Honorary Professor of Education at the North East London Polytechnic; former Chief Inspector for Primary Education in HM Inspectorate of Schools

Co-opted:

Mr T Christie  
Member of the University of Manchester Department of Education, and the Northern Examining Association

Observer:

Mr B C Arthur (HMI)

Secretariat:

Ms G G Beauchamp  
Mrs H J Field  
Mrs L F Cooke

In attendance from the Department of Education and Science:

Miss J H Bacon  
Mr M D Phipps
APPENDIX B

ELIZABETH HOUSE
YORK ROAD
LONDON SE1 7PH
01-9349000

Professor Paul Black OBE
King's College London
Chelsea Campus
Centre for Educational Studies
552 King's Road
LONDON SW10 0UA

30 October 1987

NA TIONAL CURRICULUM: TASK GROUP ON ASSESSMENT AND TESTING

1. The purpose of this letter is to offer more detailed guidance on your task than is contained in your Terms of Reference.

Relationships with the Subject Working Groups

2. You will have seen the letters of guidance which I have sent to the Chairmen of the Mathematics and Science Working Groups. I have asked those Groups to recommend attainment targets which provide objectives specific enough to offer a sound basis for the assessment and testing of pupils' progress and performance; and programmes of study detailing the content, skills and processes which all pupils need to be taught. I have made clear that the main focus of those Groups' work is to be on attainment targets and programmes of study, but I have asked them, after consultation with your Group, to offer advice in broad terms by the end of November about assessment in relation to the attainment targets they are developing, and particularly what might appropriately be measured by nationally-prescribed tests. I have also asked them to offer general advice in their final reports (June 1988) about assessment, including testing, in relation to their subjects, which should take account of the overall strategy to which I hope your recommendations will lead.

3. I shall be sending broadly similar letters to other Working Groups for foundation subjects when they are set up. Your Group will need to liaise closely with the operational Subject Working Groups, and with the Committee on English chaired by Sir John Kingman. The process of developing attainment targets, programmes of study and an overall system of
assessment, including testing, will be iterative.

Scope of the task

4. I seek advice on a coherent system of assessment, including testing, to cover the whole period of compulsory schooling. I hope that your Report will include general criteria applicable to the system as a whole but capable of accommodating appropriate variation e.g. between subjects; and more specific criteria for the choice and use of assessment and testing arrangements for particular subject areas or age-ranges. The latter in particular should take account of discussion with the operational Subject Working Groups and with informed people and organisations in other subject areas.

5. The Group's recommendations must be practicable to implement and cost-effective. I hope that your recommendations will take account of the very considerable amount of assessment which is already carried out as a normal part of teaching and learning in our schools, and will recognise that all forms of assessment affect the teaching and learning assessed. I am looking for arrangements which, by supplementing the normal assessments made by teachers in the classroom with simply-administered tests, will offer a clear picture of how pupils, individually and collectively, are faring at each of the age points. Such arrangements should help to promote good teaching.

The purposes and nature of assessment

6. Our consultation document recognises a number of purposes which assessment, including testing, may fulfil, and different uses to which information derived from assessments may be put. These include diagnostic or formative purposes, mainly concerned with ascertaining what stage a pupil has reached, identifying strengths and weaknesses and planning the appropriate next steps in the pupil's education; summative purposes, concerned with recording in a systematic way the pupil's achievement overall, particularly compared with attainment targets for each subject; and purposes mainly concerned with publicising and evaluating the work of the education service and its various parts in the light of pupils' achievements. I attach importance to all of these, and expect your recommendations to cover all. But as I have said in my letter of guidance to the Chairmen of the Science and Maths Working Groups, the main purpose of the national assessment, including testing, in relation to agreed attainment targets for national curriculum foundation subjects will be to "show what a pupil has learnt and mastered, so as to enable teachers and parents to ensure that he or she is making adequate progress and to inform decisions about the next steps".

7. I seek advice on a coherent national system for assessment and testing in relation to agreed attainment targets that recognises the different purposes to which assessment is put and takes account of how they interrelate with and complement each other, bearing in mind the points that I set out in paragraph 5 above.

8. The Group will no doubt consider a range of assessment approaches including written, practical and oral tests, and a variety of arrangements for administering, marking and moderating assessments and reporting their results. I seek advice on how these approaches should come together within a commonly-applicable system to achieve the objectives I have spelt out above.
The amount of assessment

9. I hope that the Group's advice will cover both the overall volume and the combinations of assessment and testing methods which can reasonably be accommodated, either within the normal teaching process or alongside it, but without adverse effects upon teaching time or pupils' workloads.

Ages and stages

10. Our consultation document envisages four formal assessment points during compulsory schooling. There is already assessment at age 16 for most pupils through GCSE and other public examinations. My assumption is that GCSE, or equivalent level examinations approved against GCSE criteria, will provide the basis for assessment in the foundation subjects at 16+ for virtually all pupils, and that consequently the GCSE criteria will be revised as appropriate to reflect agreed attainment targets and programmes of study and to ensure that the examinations present appropriate challenges to pupils at all levels of ability. The Group's recommendations will be relevant for those conducting this revision of the GCSE national criteria.

11. Our consultation document proposes that other assessment points should be at age 14 – when decisions about examination courses to be pursued at age 16 are normally taken; age 11 – the most usual transition point between primary and secondary education; and around the age of 7, the precise age for the first assessment stage for each subject depending on the advice of subject working groups. I expect the Group's advice to distinguish what forms of assessment, including testing, would be most appropriate for each of these age points. I assume that the reporting of the results of assessment, including nationally prescribed testing, should take place towards the end of the school year in which the majority of pupils in a teaching group reach one of the age points, although cumulative assessment contributing to the final grading may take place throughout the stage preceding that. If the Group finds compelling arguments for ages and stages other than those proposed in order to offer a coherent system, it should so advise me. I shall also particularly value advice on how assessment and testing, and interpretation of the results at the first age point (7 or thereabouts) can properly take account of the wide variation in pupils' starting points in formal education, including their previous experience in pre-school education.

Differentiation

12. I am asking the subject working groups to recommend attainment targets for the knowledge, skills and understanding which pupils of a range of different abilities should normally be expected to achieve at the four age points; but so far as possible to avoid setting qualitatively different targets – in terms of areas of knowledge, skill or understanding – for children of different ability. My objective is that teachers should not have to prejudge what attainment targets are appropriate for a particular pupil; differentiation should be by outcome not pre-selection. I therefore seek advice on the measurement and recording at a range of different levels of positive achievement in reasonably discrete elements, within each common target area. Some elements might constitute essential performance targets for all pupils; others might allow for a series of graduated outcomes.

13. I shall particularly welcome advice on the application of such an approach to the two extremes of the performance range – the very able, and the least able.
Reporting on the results of assessment

14. I attach considerable importance to improving communication and understanding at various levels about educational objectives and performance; and information derived from assessment and testing will play a key part in that. I seek advice on the information which will be required to meet the purposes identified by the Group, taking account of the need for a sufficiently detailed profile of a pupil's performance to give a full and fair indication of attainment, but also of the need to be able to present aggregated information in ways that will be useful to a wider public than the teacher, pupil and his or her parents. I should welcome advice on procedures for ensuring consistency and compatibility between different approaches to assessment, including testing, and for aggregating results. I also seek advice on a standard scale or scales of marks for describing attainment, and would welcome the Group's views on the potential role of records of achievement in recording and reporting performance and attainment, at all stages of schooling. The Group should advise on the practical implications of the reporting and recording arrangements it proposes for schools and LEAs.

Administration and moderation arrangements

15. Our consultation document suggests that arrangements will need to be made for the development, piloting and standardisation of assessment instruments including tests; for storing and retrieving such instruments; and for distributing them to schools. I should be grateful for any advice the Group wishes to offer on such arrangements in the light of the system it recommends.

16. The consultation document also proposes that assessment including the administering of nationally prescribed tests should be done in schools by teachers; but that arrangements should be made, probably with the GCSE examining groups, for external moderation of this work. Again I should be grateful for the Group's views on cost effective and practical arrangements for securing "quality control" of school-based assessments, so as to secure credibility and confidence within the education service and with the public at large.

Conclusion

17. I look forward to receiving the Group's report by Christmas 1987. I recognise that it may not be possible to cover all the ground above by then, but hope that you will be able to advise on the main issues in paragraph 1 of your terms of reference and paragraphs 4-13 above, leaving other issues and fuller elaboration for later advice. I shall make your report, together with my conclusions, available for the guidance of the subject working groups, and intend that it should be published.
APPENDIX C

Meetings with invited groups

During the course of its work the Task Group on Assessment and Testing met with various individuals and organisations to collect evidence and discuss issues.

1. Graded Assessment

Representatives of THE JMB Staged Assessment in Literacy, London Graded Assessment in Mathematics, the Oxford Certificate of Educational Achievement, the Kent Mathematics Project, London Graded Assessment in CDT, and the London Graded Assessment in Science Project submitted materials and spoke about the work of their schemes. Discussion centred on the motivating effect of clear objectives and short term goals, credibility in terms of public recognition and confidence, and resource implications including development costs, INSET and moderation. Developing practice in graded assessment in modern languages was also considered (see Page and Hewitt "Languages step by step" CILT).

2. The Assessment of Performance Unit

Representatives of the Mathematics, Language, Foreign Languages, Science, and Design and Technology teams submitted materials and spoke about the work of their schemes. Discussion covered on specifying components within subject areas, the purposes of testing, the development and delivery of assessment tasks, the time required, the range of assessment tasks and INSET implications.

3. Chief Education Officers

Chief Education Officers representing the local education authorities of Hampshire, Lancashire, ACC, Trafford, AMA, Croydon, Northumberland, and ILEA, together with the Education Officers of the ACC and AMA submitted materials and spoke about assessment in their authorities. Discussion covered the extent to which assessment was employed, the purposes it served (whether summative, formative or diagnostic), the scope and nature of assessment and testing, resource implications and reporting results. Data on the extensive use of testing in local education authorities was also obtained, from the NFER records and from the study "Testing Children" by Gipps, Steadman, Blackstone and Stierer (Heinemann 1983).

4. Experts on and Practitioners of Assessment in Other Countries

Experts on assessment systems in other countries, together with Headteachers from Germany and France spoke about the work of their assessment schemes. The discussion centred on systems operating in France, the Federal Republic of Germany, the Netherlands, Sweden, Canada, and the USA. Issues discussed included the extent of the use made of external standardised tests, use of other tests (including diagnostic tests), the nature and scope of in-school assessment, the scale of marks or grades, the arrangements made for recording and reporting, and the relationship of assessment systems to external (qualificatory) examination systems.
5. **The Kingman Committee**

Members of the Kingman Committee spoke about their work. Issues discussed included the moderation of assessment and testing, the task-specific nature of performance in English, differentiation by outcome and by pre-selection of tasks, attainment targets, the assessment of children at the age of 7, and the need to avoid cultural, racial, linguistic and gender bias in assessment instruments.

6. **School Assessment Systems**

Head teachers of primary and secondary schools in ILEA, West Glamorgan, Cornwall, and Bedfordshire submitted materials and spoke about the work of their assessment schemes. Discussion highlighted the multiplicity of assessment arrangements within schools and between schools, the purposes of assessment, the uses to which results are put, the importance and value of moderation, particularly by teams of teachers, reporting to parents and resource implications.

7. **Science and Mathematics Working Groups**

Close contact was maintained throughout by joint meetings of the chairman with the chairmen of these two groups, and by his attendance at each group for particular discussion of TGAT’s evolving policy.
APPENDIX D

This appendix gives informal accounts of two tasks actually carried out in primary schools. Although these were not developed and carried out as standard assessment tasks, they will serve to illustrate how carefully structured class activities could provide opportunities for systematic assessment, in a range of profile components, within a context of activity that was meaningful and motivating for a primary pupil. It is recognised that extensive development would be used before these could be adequate for national assessment purposes – here they serve to illustrate in general terms the type of activity envisaged.

First example: A task for seven year olds

The following task was one of several used with an infant class of 5-7 year old children as part of a topic entitled "Winter through to Spring". The task involved estimation, observation, measuring, drawing sensible conclusions and recording, and there were opportunities at various points throughout for the assessment of profiles in Language (oral and written), Mathematics and Science (problem-solving and measurement). The children, in the main, worked in groups and arrived at conclusions collaboratively although there was ample opportunity for individual work.

First activity – Discussion, interviewing, collecting data

Language: Question asked: How do we keep warm in winter?
Discussion.

Maths: Make a survey of the clothes worn by the children in the classroom.

Maths/Art: Record this information (graph).

Second activity – Estimation, measurement of temperature and time

Language: Question asked: Which materials keep us warmest?
Wrap one layer of a different material around each of 6 milk bottles - cotton, fur, wool, felt, newspaper, nylon.

Science and Maths: Fill the bottles with warm water.
Measure the temperature in each, using a thermometer, and record it.
Estimate which bottle of water will cool first.
Measure the temperature every half-hour and record.
Which bottle of water grows cold first?
Why?
Which stayed warm?
Why?

Third activity – Estimation, measurement of temperature and time, problem solving, recording, discussion, sequencing, writing, computer work
Science: Question asked: What would help to keep the bottles of water warm longer?

Language: Discussion.

Science: Repeat the experiment, first wrapping four layers of one type of material around each bottle. Are the results different? How? Why? Which material kept the water warm longest?

Maths: Put the materials into sequence, coolest to warmest,

Health Education: How is it best to dress in cold weather?

Language: Write, or use the computer to write, the story of this experiment.

Second example: A task for eleven year olds

The following is a task that has been used in several primary schools by teachers on an INSET course designed to develop use of good materials for pupils to learn by active engagement. The activity was to involve both the hand and the brain. Discussion by pupils, in groups and as a class, was an important feature: practical tasks have the advantage of giving pupils something to talk about. The task is specified below, in outline only, as a sequence of eight activities. These were covered in about four afternoon sessions. The experience showed that the task is too difficult in this form for 7 year olds, but suitable for 11 year olds.

Materials Task

First activity – Discussion

What do the children know?

Question asked: What different materials are there in the room?

Observation (science): i. Collecting children's ideas

Language: ii. Reviewing children's words and distinguishing between 'material' words and 'object' words.

Second activity – Collecting and sorting

Children bring in different materials and sorting them.

Observation (science): Discussion: what criteria were used for sorting?

Classifying (mathematics): What were the difficulties in sorting?
Discussion/Language: "Guess the group" game.

Third activity – Representation of sorting activity
Mathematics: What different groups could selected items belong to? (Matrix diagrams: 1 x 2, 2 x 2 matrix).
Science: Children sorting by two attributes – to move them from the notion of a material to the idea of a “characteristic” of a material.
Language: 

Fourth activity – Usefulness of sorting
Why do people group things in particular ways?
Language: Children describing and sorting different woods.
Discussion: Which people need to know about wood?
What do they need to know? (Focus on characteristics of wood.)

Fifth activity – Re-sorting materials into metals and non-metals
Science: Which people need to know about metals?
Language: Discussion: children's ideas about metals – the words they use to describe them and the uses they attribute to them.
Observation (science): Focus on one use – a clip. Children observing the use in pen clips, hair grips, slides and safety pins.
Discussion: children describing the job to be done by the metal and what the metal has to be like (focus on metallic properties).

Sixth activity – Which metal is the best bender?
Science: Children grappling with what to measure, whilst carrying out a trial and error investigation.
Language: Discussion: What do we mean by best bender?

Seventh activity – Discussion of investigation
Science: Children evaluating their trial and error investigations.
Language: Focusing on: purpose for the investigation, i.e. information for pen manufacturer, type of information needed, and what has to be judged.
Mathematics:
Eighth activity – Further investigation

Mathematics: Focus on how to measure and controls.

Science:

The labels on the left indicate where opportunities would arise for assessment of various skills. The way in which this could be done in standardised form is not set out here: substantial development and trialling would be needed to establish such methods. There are several opportunities for monitoring discussions, for observing practical work by pupils working in groups, and through written work. The task could clearly be broken into two parts (after the fourth activity) but its length was not a disadvantage in practice. It would not be difficult to arrange for each assessed component to be involved in at least two points over the eight activities, thus giving more time for a teacher to assess all pupils once on that component during the whole task.

Important lessons learnt from the use of this and similar tasks are as follows:

• the conceptual understanding of the subject of a task can be a hurdle for many children: if this is not overcome their other competences cannot be fully assessed.

• topics must be so chosen that they give a wide range of possible development – this allows differentiation to be developed within the task as children show competence or difficulty at early stages, and also ensures that a range of profile components can be assessed.

• the task is both curriculum/learning material and assessment material: teachers find it very hard at first to keep both aspects in balance.

• in particular there is a temptation to close down open discussions and over-define situations or outcomes, partly because of a desire to "teach" and partly because of a belief that the assessment context must be rigidly defined.

• the use of group work, encouraging discussion within and reporting by groups, and of genuine class discussion, exploring children's ideas and results and not inserting teachers' own too early, is an unusual style for many, who find it hard to adapt.

• teachers need more than initial instruction – they also need mutual support through meetings, en route during the work, to exchange experiences and give mutual comfort.

These tasks were not design to cover (what might become) the full range of profile components at the primary ages. For example, opportunities to use artistic or craft skills might be taken up if they were re-designed for this broader purpose.
APPENDIX E

Test and assessment methods: examples of the variety available

The intention behind this appendix is to provide a selection of forms of assessment and types of test in order to illustrate the range and variety of testing techniques which are available or are capable of development. It should not be seen as a full compendium of tests and test types, but rather as a potpourri of possibilities. Two further examples are included in Appendix D: these are set apart because they illustrate a particular approach recommended for testing in the primary phase.

As used in this appendix and elsewhere in the report, a test is a formal procedure for evaluating aspects of an individual's knowledge, skills and understanding – or 'behaviour'. It provides a means of eliciting samples of behaviour in a standardised manner and prescribed procedures for grading it. The breadth of this definition indicates the great variety of approaches which are encompassed by tests. They differ in their modes of presentation, their modes of working and in the modes of responding. Modes of presentation can include oral questions, written questions, video films, computer displays and demonstrations. Modes of working can be mental, written, oral, practical and combinations of these. The responses, or products, of the test can be in written or oral form or require active performance, constructions or combination of these forms.

Tests differ also in the size of the samples of behaviour they assess. Some tests are combinations of single short questions; others have tasks which require the integration of several skills and a variety of aspects of knowledge. Both these approaches have advantages and disadvantages which make them more or less appropriate in specific situations.

Some of the contrasts between tests composed of discrete questions and those consisting of integrated tasks are shown in Table 1. Essentially, tests consisting of a series of short questions can sample more widely and have greater control over standardisation; integrated tasks provide more realistic samples of behaviour and may therefore provide more valid measures of attainment particularly for younger pupils. In the integrated tasks, the emphasis on formality is reduced and, for example, interactions with assessors may be included some of it as part of the assessors work of gathering results. This interaction is, though, still subject to control and a set of rules would still govern the information imparted in the exchanges.

The choice of a particular test or testing technique depends on several factors. Primary among these is the purpose for which the results will be used but practicalities can also be important. These would include the time available, the resources available, the age and other characteristics of the pupils, the training requirements and the manner in which the results are to be processed and reported.

For each of the examples that follow, the source, i.e. the test collection or agency from which it has been selected, is given and discussed; this is followed by a specification of the example, followed by the example itself. For some of the integrated assessments, the complete task is described, for others only a summary is given.
Material from some of the newer initiatives in assessment is also included: the Oxford Certificate of Educational Achievement (OCEA); and the various London graded assessments projects: GAIM, (Graded Assessments in Mathematics); GASP (Graded Assessments in Science Project); and GAML (Graded Assessments in Modern Languages, London).

The OCEA material is given as an example of the records of achievement initiative, being one of the nine pilot schemes supported by the Department of Education and Science. The aim is to assess a wide range of pupils' achievements and experiences. Students participate actively in the learning process through the establishment of specific negotiated short-term goals, which tends to improve motivations. Only a flavour of the many possible assessments can be given, in three subject areas (Examples 17 to 19).

The graded assessment projects are complete assessment schemes in which the pupils must meet specific criteria related to different levels of attainment. Pupils are assessed only when they are ready to tackle the range of tasks at a particular level. There is an emphasis on setting tasks which are likely to be relevant to the pupil's own experience and to have practical applications and value to adult and working life. This leads to a great variety of assessment tasks of which only one is included here (Example 20). The final illustration, Example 21, is given as a possible task which could provide assessments of science, language and mathematics on the basis of the same piece of work undertaken by an 11 year old pupil.

<table>
<thead>
<tr>
<th>Tests composed of discrete questions</th>
<th>Tests consisting of integrated tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can sample from a wide variety of knowledge, skills and understanding.</td>
<td>A single task may be limited to a few aspects of knowledge, skills, and understanding but may offer detailed assessment of these.</td>
</tr>
<tr>
<td>Can make it difficult to assess the integration of knowledge and skills.</td>
<td>Can allow the integration of knowledge and skills.</td>
</tr>
<tr>
<td>Can be artificial.</td>
<td>Can be realistic.</td>
</tr>
<tr>
<td>Can be easy to administer.</td>
<td>Administration in a standard form may need some training.</td>
</tr>
<tr>
<td>Can be administered in a relatively short period of time.</td>
<td>Can be time consuming.</td>
</tr>
<tr>
<td>Marking can be relatively objective given suitable marking schemes and training.</td>
<td>Marking will be relatively subjective unless procedures are carefully set out and some training given.</td>
</tr>
<tr>
<td>Marking can be accomplished relatively easily.</td>
<td>Marking can be difficult and may need special training.</td>
</tr>
<tr>
<td>Readily yields numerical data which can be difficult to interpret meaningfully.</td>
<td>Qualitative data can be diagnostically illuminative and are easily gathered. Analysis of responses can yield categories of response and numbers of pupils in each category.</td>
</tr>
</tbody>
</table>

Table 1
Example 1

Source

Early Mathematics Diagnostic Kit
(Lumb and Lumb, 1986, NFER-Nelson)

This is a collection of test items designed to identify early difficulties in mathematics learning. Items consist of pictures and a question read aloud by the teacher. Some items also allow the use of apparatus such as Unifix cubes. The items are related to specific behavioural objectives and are linked to suitable remedial activities.

Specification

Subject area: Mathematics
Age group: 4-8 years
Individual administration

Presentation mode: Oral (teacher) and pictorial
Task mode: Mental
Response mode: Oral and written

Individual questions not aggregated.

Example

How many? (read out by teacher)
Note the highest number of objects the child can identify correctly without verbally counting. Give child paper and pencil.

Write down the number. (read out by teacher)
Note how the child writes the number. Is the number formed fluently and quickly?
Example 2

Source

Maths 7
(NFER, 1987, NFER-Nelson)

This is a test designed to assess the mathematics attainment of children near the end of the school year in which they reach their seventh birthday. It is principally a broad monitoring test.

Specification

Subject area: Mathematics
Age group: 6-7 years
Group administration

Presentation mode: Oral (teacher) and pictorial
Task mode: Mental and/or written
Response mode: Written

Individual questions aggregated into total score.

Example

(Teacher reads) how much money is in the piggy-bank? Write down the answer on the line.
Example 3

Source

Maths 11
(NFER, 1984, NFER-Nelson)

This is a test designed to assess the mathematics attainment of children towards the end of the academic year in which they reach their 11th birthday. It is principally a broad monitoring test which is norm-referenced. Procedures are given for teachers to compare this performance of their own groups with the national facility value for each question.

Specification

Subject area: Mathematics
Age group: 10-11 years
Group administration

Presentation mode: Written
Task mode: Mental and/or written
Response mode: Written

Example

The graph shows the weight of a baby girl in the first eight weeks of her life:

48. What was the baby's weight at birth?
   Answer _______ kilograms

49. How heavy was the baby when she was 8 weeks old?
   Answer _______ kilograms

50. How old was the baby at the end of the week in which she gained most weight?
   Answer _______ weeks old
Example 4

Source

Special Access Mathematics 1A
(NFER, 1986, NFER-Nelson)

This is a multiple-choice test restricted in use to LEAs who use it for monitoring exercises. It is norm-referenced and consists of discrete questions designed to test the understanding of mathematical concepts and processes or to test the use of mathematics in everyday situations. The distracters (wrong options) were selected to reflect procedural errors commonly followed by children of this age and misconceptions often held by them; that is, they provide diagnostic indications.

Specification

Subject area: Mathematics
Age group: 10-11 years
Group administration

Presentation mode: Written
Task mode: Mental and/or written
Response mode: Multiple-choice

Example

This pie chart shows the favourite TV channel of 1600 people:

How many people like Channel 4 most?

A. 45  B. 200  C. 320  D. 400  E. 90
Example 5

Source

Chelsea Diagnostic Mathematics Tests
(Hart, Brown, Kerslake, Kuchemann, Ruddock, 1985, NFER-Nelson)

These are a series of tests, each concentrating on one topic within the secondary school mathematics curriculum. They were designed as diagnostic instruments to be used for ascertaining levels of understanding and for identifying the incidence of errors. The item shown is from the measurement test.

Specification

Subject area: Mathematics
Age group: 11+ to 14+
Group or individual administration

Presentation mode: Written
Task mode: Mental and/or constructions (depending on child)
Response mode: Construction

Example

On the squared paper below, draw:

(a) A square whose area is twice (double) the area of square S. Put a cross on (a) if you think it is impossible.

(b) A square whose perimeter is twice (double) the perimeter of square S. Put a cross on (b) if you think it is impossible.
Example 6

Source

Item from tests developed for Graduated Tests in Mathematics for the Lower Attainers in Secondary School Feasibility Study.

This project developed a series of tests for use with the bottom 40% of pupils in secondary schools. Assessment materials in a variety of modes were investigated, including written, oral, practical and investigative modes. One task investigated was a realistic simulation of shopping in order to test the Cockroft Foundation List target "Recognise coins and notes, and know that 100p = £1. Handle money with confidence".

Specification

Subject area: Mathematics
Age group: 13-15 years (low attainers)
Individual administration with assessor

Presentation mode: Oral
Task mode: Mental
Response mode: Oral and performance

Example

The pupil is given £7 (a £5 note and two £1 coins) and is required to buy a series of articles, check the change and state how much money is left. After some purchases the wrong change was given by the assessor and the pupil's reactions to this noted.
Example 7

Source

Maths with the Micro Survey: APU 1987

This is a test developed for the APU 1987 secondary mathematics survey. The pupil has to plan a weekly budget using a Spreadsheet package. The pupil has to adjust the budget to fit given amounts of pay using the Spreadsheet. The item gives opportunities to examine understanding of percentages and of changes in percentages given a new base quantity. The microcomputer allows amendments and corrections to be made flexibly and the realistic situation provides a vehicle in which to examine the application of mathematics to everyday life.

Specification

Subject area: Mathematics
Age group: 15-16 years
Individual administration via assessor and microcomputer

Presentation mode: Oral and microcomputer
Task mode: Mental and keyboard
Response mode: Oral and microcomputer

Example
(see source explanation above)

Screen layout during a typical run.

<table>
<thead>
<tr>
<th>BUDGET</th>
<th>PAY</th>
<th>NEW PAY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28.00</td>
<td>50.00</td>
</tr>
<tr>
<td>GUESS %</td>
<td>20.00</td>
<td>16.00</td>
</tr>
<tr>
<td>REAL %</td>
<td>17.86</td>
<td>14.29</td>
</tr>
<tr>
<td>COST</td>
<td>5.00</td>
<td>4.00</td>
</tr>
<tr>
<td>NEEDS</td>
<td>FARES</td>
<td>LUNCH</td>
</tr>
<tr>
<td>NEW COST</td>
<td>5.00</td>
<td>4.00</td>
</tr>
<tr>
<td>GUESS %</td>
<td>SAME</td>
<td>SAME</td>
</tr>
<tr>
<td>REAL %</td>
<td>10.00</td>
<td>8.00</td>
</tr>
<tr>
<td>CHANGE %</td>
<td>-7.86</td>
<td>-6.29</td>
</tr>
<tr>
<td>TOTAL</td>
<td>112.00</td>
<td>100.00</td>
</tr>
<tr>
<td>SAME</td>
<td>100.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Example 8

Source

Diagnostic Test in Addition and Subtraction
(NFER, 1987, unpublished)

This test uses the micro-computer as a diagnostic tool to build a picture of the misconceptions a child has when attempting certain arithmetic procedures. Questions are generated by the program to test hypotheses and determine which particular misconception(s) the child is following. The results are reported as an error category rather than as a score.

Specification

Subject area: Mathematics
Age group: 7-11 (children having difficulty with addition or subtraction)
Individual administration via microcomputer display

Presentation mode: Microcomputer display
Task mode: Written
Response mode: Keyboard entry

Example

Two of the largest set of questions involved are follows in the form of the actual computer print out. The pupil response has been entered in, and the comment on the right is a diagnostic interpretation produced by the programme from the evidence of the pupil response.

Questions and answers:

Question 1:
970
− 518
468  Smaller from larger
O − N = N

Question 2:
804
− 727
100  No decomposition/zero answer

After a test in which a larger number of such items are attempted by any one pupil, the programme gives a diagnostic summary for the teacher. An example is produced below. The asterisks indicate the most significant features.
<table>
<thead>
<tr>
<th>FREQUENCY OF OCCURRENCE</th>
<th>ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>*** 7</td>
<td>Smaller from larger</td>
</tr>
<tr>
<td>* 2</td>
<td>Number bonds</td>
</tr>
<tr>
<td>** 7</td>
<td>No decomposition/zero answer</td>
</tr>
<tr>
<td>1</td>
<td>O – N = N</td>
</tr>
<tr>
<td>2</td>
<td>O – N = O</td>
</tr>
</tbody>
</table>

Incorrect answers are probably due to the following error(s):

- Smaller from larger
- No decomposition/zero answer
Example 9

Source

Integrated task devised for 7 year olds by APU Mathematics team (untried)

Specification

Subject areas: Mathematics and language
Age group: 7 year olds
Individual administration or small groups

Presentation mode: Oral
Task mode: Practical and mental
Response mode: Oral and written

Example

1. Apparatus

- 200g bag of red kidney beans
- Balance
- Containers of different sizes
- Measuring jar or long cylindrical container
- Calculator

2. Task

i. Estimate the number of beans in a 200g bag of red kidney beans.
ii. Using any of the apparatus provided make a better estimate without counting out all the beans.
iii. Check using a different method from ii.
iv. Pupils to record results and to give an oral account of what they did to obtain their results.

3. Possible methods of tackling the task

i. Overall strategy:
Count the number of beans in a small unit; relate the smaller unit to the whole.

ii. Possible tactics:
   a. Use balance successively to halve number of beans until a manageable number of beans is obtained; multiply up to obtain number in 200g.
   b. Count the number of beans in a small container; find the number of small containers which will hold the 200g of beans.
   c. Pour the 200g of beans into the long cylindrical container. Measure the height; now pour out most of the beans and measure the height of a proportion of them; relate the smaller height to that of the 200g.
4. Assessment

i. Overall strategy:
   Could be assessed on a rating scale. Evidence for pupils' appreciation of the strategy would be obtained from their recording and oral account of what they did. Levels of mathematical argument used could be similarly assessed.

ii. Carrying out the task – use of mathematical skills:
   a. Accuracy of counting
   b. Accuracy of calculation
   c. Correct choice of operation
   d. Understanding and use of balance
   e. Calculator skills: accuracy of keying, order of operations, interpreting display

iii. Communication skills:
   a. Recording
   b. Oral
Example 10

Source

NFER Reading Ability Series Level A (NFER, in press, NFER-Nelson)

This is part of a new series of reading tests which assess reading as a purposeful activity using realistic material which might be encountered within or outside the classroom. Different types of text both narrative and expository are included and the children's response to these assessed through cloze passages and open-ended questions.

Specification

Subject area: Language
Age group: 7-8 year olds
Group administration

Presentation mode: Written
Task mode: Written
Response mode: Written

Example

Exercise 1
Please look at this notice, and then answer the five questions.

3. What colour are they?

4. What would you get for finding both kittens?

5. Suppose that you found the kittens, in what two ways could you let Mrs. Blake know?

1. How many kittens were lost?

2. Who lost the kittens?
Example 11

Source

Assessment of Speaking and Listening (Gracy)
Taken from APU surveys, 1982-83.

These listening and speaking tasks were administered by an assessor working with pairs or groups of pupils seated round a table. In this particular task, two pupils listen to a taped description of how a spider builds a web; then, using a sequence of pictures, they are required to tell another pair of pupils about the process.

Specification

Subject area: Language
Age group: 11 year olds
Individual or small group administration

Presentation mode: Aural (tape) and pictures
Task mode: Oral
Response mode: Oral

Example


**Example 12**

**Source**

*Assessment of Speaking and Reading*
Task taken from APU age 11 oracy survey (1983)

**Specification**

- **Subject area:** Language
- **Age group:** 11 year olds
- **Administration:** Small group
- **Presentation mode:** Reading and aural
- **Task mode:** Oral
- **Response mode:** Oral and written

**Example**

This is a task for two children and involves three basic activities for assessment.

These are:

1. Extracting information from a simple text.
2. Exchanging that information.
3. Reaching consensus through collaborative discussion.

Each child is given information about one of two fictitious plants (the Kava and the Bulo) presented in comic strip style. The children are asked to read the information and note down the merits and demerits of their allotted plant with respect to its suitability for planting on an island which has a range of given agricultural problems.

Having completed this task, they then proceed to tell their partner what they have found out about their plant. This exchange leads on to discussion about the advantages and disadvantages of each and a final decision about which should be cultivated on the island.
The soil has to be dug well, to grow the Kava plant.

A lot of compost has to be added to the soil to feed the plants.

Once the plants have started to grow, they need very little water.

The plants grow stronger if the weeds are not allowed to flourish.

The Kavas are resistant to disease and pests.

When ripe the Kavas are difficult to harvest, as the leaves are very prickly.

They need a lot of preparation to cook them.

They are extremely tasty vegetables.

The seeds are very cheap.

The seeds have to be planted every year.
Example 13

Source

Assessment of Writing
Tasks taken from APU surveys (1979-83)

The survey framework included extended pieces of writing with relevance to different curricular areas, as well as shorter writing tasks such as editing text produced by pupils of the same age, commenting on reading passages, and making notes for specific purposes. Other tasks required the integration of writing with diagrams or illustrations, as in the preparation of posters.

The writing is assessed using both impression and analytic scales, applied by teachers of the relevant age groups.

Specification

Subject areas: Language and writing
Age group: 11 year olds
Group administration

Presentation mode: Written and oral
Task mode: Written
Response mode: Written

Example

THE BALLOON GAME

A group of people who all think they are very important to the rest of us are riding in a balloon. They are a farmer, a cook, a teacher, a social worker, a prime minister, a nurse, an actress and a miner.

The trouble is that some of them have to jump out or the balloon will not fly high enough. Whoever can argue the most convincingly will stay longest in the balloon. Decide who you think is really the most important person of the eight and write a speech for him/her which will convince the others that he/she can stay in the balloon.

The purpose of this task is to assess pupils' ability to write persuasively and from a particular point of view.

THE EARLIEST THING I CAN REMEMBER

Think of the earliest thing you can remember when you were very young. It could be a place, a person, something that happened, even an object. Describe it as clearly as you can.

The purpose of this task is to assess pupils' ability to produce a short piece of descriptive writing.
**Example 14**

**Source**

*Assessment of Speaking and Listening (Oracy)*  
Task taken from APU age 11 oracy survey (1983)

**Specification**

Subject areas: Language and science  
Age group: 11 year olds  
Administration: Part individual, part small group  
Presentation mode: Aural and practical (instructions from assessor to be followed by pupil)  
Task mode: Oral  
Response mode: Oral and written

**Example**

This task comprises a sequence of activities based on a scientific experiment similar in principle to "Newton's cradle". When a marble runs down a plastic track and hits one end of a stationary group of four, the marble at the other end of the group is impelled up the track, rises, slows and stops, and then runs back to repeat the whole process (see diagram):

1. The assessor instructs pupil A on how to carry out the experiment and to observe what happens.  
2. Then pupil A similarly instructs pupil B.  
3. Pupil B describes his/her observations to the assessor.  
4. Both pupils discuss the experiment and develop a hypothesis to explain their observations.  
5. Pupil A states this hypothesis to the assessor.  
6. Two more pupils are brought in, and produce a written account of the experiment, with diagrams.

**Assessment**

Pupil A: Instructing Pupil B (2), hypothesising (4), reporting the hypothesis (5) and the written outcome (7).  

Pupil B: Describing observations (3), hypothesising (4), informing written outcome (7).
Example 15

Source

Task taken from APU German Test

The APU Modern Languages framework for assessment involves tests of listening, speaking, reading and writing. All are set in realistic contexts and the pupils are asked to carry out tasks which would be appropriate within those situations. An example of an item in a listening test is illustrated below.

Specification

Subject area: Foreign language
Age group: 13 year old learners of German as a first foreign language.

Presentation mode: Oral
Task mode: Oral
Response mode: Selection of correct picture.

Example

You are talking to another camper. She tells you what her camping holiday has been like.

Example 16

Source

Task from Piagetian Reasoning Tasks (previously CSMS Science Reasoning Tasks) (Wylam, H. and Shayer, M. 1978)

These tests are designed to assess the ability of children and adults to use concrete and formal reasoning strategies. The tasks were intended to be used as an aid to curriculum development in science teaching. In the task shown below the pupil is asked to say how he would plan tests to distinguish the effects of three variables: how the length, weight and push of a pendulum determine the period of oscillation. Only the length is important, but the student must design an experiment controlling the appropriate variables in order to reach this conclusion.

Specification

Subject area: Science
Age group: 9 to 16 years
Individual administration

Presentation mode: Written and diagrammatic
Task mode: Practical
Response mode: Written

Example (see source above)

Equipment:

- Stopclock or watch with second hand
- 2 weights (say 100 and 400 gram slotted weights on hangers)
- 2 strings looped at either end. (Loop to loop distances of 69 cm and 35 cm will give about 17 and 22 swings in half a minute. The numbers are not critical, but try to avoid lengths giving either 15 or 20 swings.)
- Firm support to hang pendulum from.

For information on the use, development, statistics etc of this task see Science Reasoning Tasks Booklet (NFER, Windsor).
Example 17

Source

OCEA Mathematics

The OCEA teacher’s guide to mathematics, "Putting it into Practice", gives 57 sets of assessment materials, together with illustrative solutions. The materials are divided into short problems, investigations, practical problems, quick response questions and projects. The example shown is one of the practical problems.

Specification

Subject area: Mathematics
Age group: 13-15
Individual task

Presentation mode: Written with illustration
Task mode: Practical
Response mode: Construction

Example

BIRD CAGE COVER

Sketch a free-hand plan for a cover for this bird cage.

Choose a scale to fit your paper and draw a scale plan for the cover.

Use your plan to make a model of the cover in paper, to see if it works.
Example 18

Source

OCEA Modern Languages

The OCEA Modern Languages assessment framework is built on the four skills of listening, speaking, reading and writing. Detailed prescriptions are given for the assessment of these skills and combinations of them. These are largely realistic simulations of relevant situations. One of these, to assess talking, is shown below.

Specification

Subject area: Modern languages – French
Age group: 13 -15
Small group task

Presentation mode: Written
Task mode: Oral
Response mode: Oral

Example

Topic: Shopping for a present
Context: Student A/teacher – you are a shop assistant in a French gift shop. Student B – you are in a gift shop in France buying a present for a friend or relative.

Aim of task: Through exchange of information Student B buys a gift which he/she can afford for a friend or relative.

<table>
<thead>
<tr>
<th>POUPEES</th>
<th>POSTERS</th>
<th>T-SHIRTS</th>
<th>A. You are a shop assistant in a French gift shop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costume breton 29F 00</td>
<td>Grands 15F 00</td>
<td>Avec image 23F 00</td>
<td>[STUDENT/TEACHER ROLE]</td>
</tr>
<tr>
<td>Costume normand 30F 00</td>
<td>Petits 8F 00</td>
<td>Avec motif 25F 50</td>
<td></td>
</tr>
</tbody>
</table>

Elle collectionne les poupées

11 ans

28 francs

You are in a gift shop in France buying a present for a friend or relative.
Example 19

Source

OCEA Science

The OCEA Science assessment framework is based on four processes: planning, performing, interpreting and communicating. Each of these is linked to a small set of skills; for example, planning is linked to (1) producing testable ideas and (2) designing investigations.

These are then further divided into criteria which are positive statements describing achievements within the skills. In all there are 34 criteria and any activity can be used to assess a variety of these. The student activity shown below could be used to assess one skill associated with producing testable ideas, that the pupil can suggest other ideas related to those already investigated, but it could also be used for many other skills.

Specification

Subject area: Science
Age group: 11-14 years
Individual task

Presentation mode: Written
Task mode: Practical and written
Response mode: Practical and written

Example

Worksheet: Footpath

Many students take a short cut across the grass to get to the sports hall.

Map:

Balbir said that the grass would be thinner where people had been walking. She thought of a way of finding out if this was true.

1. Think of something else that might be different about the path and write it on your sheet.
2. Think of an experiment to see if you are right and describe it on your sheet.
3. Do your experiment and then write about it.
**Example 20**

*Source*

The Graduated Assessments in Mathematics scheme ILEA, WLSEB and King's College London is similar to OCEA's in that it provides a categorisation of topic criteria which can be assessed in many different ways. The pupils' progress on particular criteria can be assessed through classwork, self-assessment or through review sheets which may be undertaken as a final check after all the topic criteria in one area have been recorded as satisfied. Some assessments arise from group activities concerned with a practical problem. One of these is shown below.

*Specification*

- **Subject area:** Mathematics
- **Age group:** Age 11 upwards
- **Age 2-3 students in group or independently**
- **Presentation mode:** Written
- **Task mode:** Oral and written
- **Response mode:** Written, diagramatic

*Example*

**Breakfast Cereals**

1. What is the most popular breakfast cereal in your class?
2. Which cereals give best value for money?

Your group must decide *how* to find the answers to these two questions.

- Do you need more information?
- How will you find it out?
- What jobs need to be done?
- Who will do what?

Discuss and decide how you will present your results.
**Example 21**

*Source*

SAILBOATS

Integrated tasks devised for eleven-year-olds by APU teams (untrialled in this version - used in an earlier version in science monitoring work). This task is rather similar to those in Appendix E and could be used as part of the collection in which they might (after full development) be made available.

**Specification**

Subject area: Science, mathematics and language
Age.group: 11 year-olds
Small group administration

Presentation mode: Oral and written
Task mode: Practical, oral, written and mental
Response mode: Oral and written

**Example**

Aim of the experiment:

To discover the most effective shape of boat given:

1. A choice of two boats;
2. Two different types of sail;
3. A third sail of given shape but choice of size to be decided by the pupil.

Apparatus:

- Tray 60cm x 40cm x 3cm
- 2 different shapes of "boats"
- 2 split pin type masts
- Minifan with stand
- Metre rule
- Stop clock (marked in one-tenth second divisions)
- Small bucket
- Plastic sheet

The equipment should be laid out on the plastic sheet as illustrated below:
Administration

The task involves collaboration between three pupils. The pupils are told that the problem involves finding the best shape of sail for some model boats. The assessor shows the pupils how to fit sails to the model boats and how to use the other equipment. The pupils are told that they are to record how long the boats take to sail the length of the tray or part of it. The assessor then ascertains that the pupils understand the task.

One of the three pupils is given charge of the experiment. The pupil in charge is also in control of the fan and release of boats. The other two have, respectively, the tasks of timing the boats with the stop-clock and recording the experiment.

Two sets of sails are provided of different sizes.

1. Each of the pupils is asked to make drawings of the sails and of the boats and to record the dimensions on the drawings (including the height).

2. They are told to find out the area of the sails and boats.

3. The three pupils are then asked to discuss how they will carry out the investigation, and

4. to record on their work sheet the steps they have agreed to take in the experiments.

5. The pupils are then asked to discuss which boat and sail they think will give the fastest speed. They are not at this stage prompted to identify what factors relating to the size and shape of the boats and sails are important in determining this.

6. If after ten minutes pupils have not referred to any of the measurements of the boat in their discussion they are prompted to consider taking the following ones into account:
   a. The dimensions/shape of sails
   b. The dimensions/shape of boats
   c. The height of the sail
   d. The area of the sail.

7. The pupils are then told to carry out the series of experiments and that each of them in turn
will act as time-keeper and as recorder.

Reporting Back

8. The pupils are asked to complete a report sheet giving their results and to comment on how their findings compared with their predictions.

Classroom Follow-up

Based on their findings all pupils in the class could investigate sailing ships and discuss their suitability for various applications, speed and manoeuvrability. They could also carry out further experiments investigating the effects on speed of varying the dimensions of the sails.

Assessment

• Ratings of effectiveness of group discussion and oral presentation (hypothesising, predicting, explaining) (3, 5).

• Assessment of the clarity and appropriateness of features of the written records (sequencing, tabulation, recursivity) (4, 8), plus:
  o checklist of carrying out of mathematical operations (1, 2)
  o checklist of carrying out scientific operations (5, 6, 7)
Acknowledgements

Acknowledgements and thanks are due to the following organisations for permission to use their material.

- National Foundation for Educational Research
- NFER-Nelson Publishing Company
- Oxford University Delegacy of Local Examinations for OCEA material
- University of London School Examinations Board for Graded Assessment material
- Science Reasoning, King’s College, University of London
- Assessment of Performance Unit
APPENDIX F

NATIONAL TESTING AND EQUAL OPPORTUNITIES

Submission by the Equal Opportunities Commission to the DES Task Group on National Testing

December 1987
# Table of Contents

1. Summary 1

2. Introduction 3

3. Evidence of Gender Differences 3
   3.1 Language Differences 3
       3.1.1 Reading 3
       3.1.2 Writing 4
       3.1.3 Test Format 4
       3.1.4 Motivation and Context 4
   3.2 Mathematics Differences 4
   3.3 Verbal and Non-Verbal Ability Differences 5
   3.4 Science Differences 5
   3.5 Summary 6

   4.1 The Use of Test Results 7
   4.2 Adjusting for Gender Differences 8
   4.3 Gender-Fair Tests 9

5. Further Work 9

6. References 11
1. **Summary**

1. There are established differences in test performance between girls and boys throughout the age range 7 to 16 and across different areas of the curriculum. Differences in performance at various ages have been shown in language, mathematics, science and verbal and non-verbal reasoning (section 3).

2. Various aspects of test design have differential effects on girls’ and boys’ performance. The use of practical tests, the form of questions adopted (eg. multiple choice vs. essay questions) and the context and content of individual test items have all been shown to affect the relative performance of girls and boys (3.1.3, 3.1.4, 3.2 and 3.4).

3. These findings have important implications for any proposed national system of testing pupils at 7, 11 14 and 16:

   (i) Where the results of a number of tests are combined to give a single overall test score the resulting gender difference will reflect the weighting given to the component test areas (3.5).

   (ii) If test results are to be used to compare schools, it is important that the proportion of pupils of each sex in the relevant age group within each school should be known and taken into account. The effects of gender differences on inter-school comparison will, of course, be greatest where the comparison involves a single-sex school (4.1).

   (iii) If test results are used to allocate individual pupils to ability groups for particular subjects, differences in test performance will affect the gender composition of different groups. If scores from a number of tests are combined and used for general banding or streaming purposes, the gender balance of each band or stream will be dependent on the weighting given to the different component tests (see 3(i) above) (4.1).

   (iv) Tests can be constructed to reflect any desired degree of gender difference, or none. There are no generally accepted objective external criteria for test construction which can guide a test constructor in choosing the relative weightings for different test questions or topic areas. If tests are constructed, on grounds of equity, to give equal average scores for girls and boys, consistency would require that this should be done at all ages. This would raise questions about other assessment procedures such as GCSE (4.1 and 4.3).

   (v) An alternative to eliminating gender differences in average test performance by revising the construction of the tests is to standardise test scores according to different gender norms. In the EOC’s view this would constitute unlawful discrimination (4.2).
2. **Introduction**

This report is divided into two parts. The first gives a brief summary and evaluation of the evidence for gender differences and the second part looks at some possible consequences of these in a system of national 'benchmark' tests. The areas studied are those of language, mathematics, ability tests and science.

Because of time limitations, only a very general coverage of the issues is possible. Nevertheless, because the subject is complex, touching upon basic issues of educational achievement, a more extensive treatment is desirable. In a final section an outline is given for such further work.

3. **Evidence for Gender Differences**

3.1 **Language Differences**

3.1.1 **Reading**

In terms of vocabulary, boys show greater knowledge in Primary school (Douglas et al., 1968) with a persisting difference through the secondary school years (Wittig and Peterson, 1979).

In reading comprehension, an early (7-8 years) advantage in favour of girls, becomes very small by the age of 11 years and by the age of 16 years there is a small advantage to boys (Fogelman et al, 1978; Davie et al, 1972; Douglas et al, 1968; NCDS, 1972).

The American National Assessment of Educational Progress (NAEP, 1986a) reports superior overall performance for girls in reading up to 17 years and the British Assessment of Performance Unit (APU, 1982) does so at ages 11 and 15 years.

3.1.2 **Writing**

According to NAEP (1986b) girls do better than boys overall at all ages on writing tasks, and this finding is supported by the APU results at 11 and 15 years. NAEP divides writing into 'informative', 'persuasive' and 'imaginative.' In all types of writing girls perform better than boys at all ages from 9 to 17 years.

3.1.3 **Test Format**

There is considerable evidence that, relatively, boys do better than girls in multiple-choice questions when compared to free response or essay questions. Murphy (1980), dealing with geography O' level exams, finds an increased pass rate for boys when the percentage of multiple-choice questions is increased. Wood (1978) reinforces this for language examinations.
3.1.4 Motivation and Context

Wood (1978) provides evidence, based on O' level English language exams, that boys do better than girls when the topic of a question is a ‘masculine’ one (e.g. trains) and girls do better when it is a feminine topic (e.g. a story about a young girl's feelings).

3.2 Mathematics Differences

At the end of primary schooling, boys and girls appear to have similar performances on, largely arithmetical, mathematics tests. By 15 years the boys show an advantage (Douglas et al., 1968). Wood (1976) shows that boys do better than girls in 15 years old public examinations in questions concerning measurement and spatial topics. The APU data (APU, 1986) confirms this at age 15 and also shows that there are only very small differences in number skills and modern algebra. Boys also tend to do better in applied and practical mathematics. At the age of 11 years boys are more confident in measurement and practical tasks than girls, and at this age the girls performed better than the boys only on computation tasks. There was only a very small difference at this age again in modern algebra. There is also a suggestion that there is only a very small difference at 11 years in problem solving tasks.

Plake et al (1982), using university students, found that the arrangement of items in a mathematics test was important. With a traditional easy-hard ordering males did better than females but for an arrangement where easy and hard items were uniformly spread across the test these differences disappeared. There appears to be little data concerning such an effect in school age pupils.

3.3 Verbal and Non-Verbal Ability Differences

At the end of primary school, girls appear to score higher than boys on non-verbal tests and considerably higher than boys on verbal tests. By the end of secondary school, boys score somewhat higher on the non-verbal tests (NCDS 1972; Douglas et al, 1968). Within the nonverbal domain, boys appear to begin to outperform the girls from about age 13 (Macoby and Jacklin, 1974).

3.4 Science Differences

The APU science monitoring programme (APU, 1982) shows that at the age of 11 years boys tend to do better than girls in the application of taught science concepts and in practical investigations. At age 15, results have to be interpreted with care since boys generally have had more exposure to science. The boys do better than girls at this age in the use of equipment, interpreting data, reading information and applying physics concepts. The latter difference is also apparent at age 13.

3.5 Summary

An important feature of gender differences is that context, content and format of test questions can affect these differences, even to the extent of reversing a difference on otherwise similar questions. The multiple-choice format appears to favour boys as does practical testing.
4. National 'Benchmark' Testing

It is proposed by the Government that nationally prescribed tests will be administered and marked by teachers at the ages of 7, 11, 14 and 16 years in the areas of mathematics, language and science. It is possible that these ages could be varied somewhat, but that would not affect the substance of what follows. At the time of writing there are no details of how the test scores will be used, but some general indications are available.

The national curriculum consultation document refers to the use of the tests for providing parents with information for comparing schools, and since it is intended that parents will be given details of their children's test results, it is likely that these results will also affect school decisions on ability grouping etc. There have been suggestions that the test results will be useful as 'diagnostic' instruments for the teacher and pupils. It is unlikely, however, that the tests will be detailed or sensitive enough for this purpose and most professionals would not accept that a single instrument properly can combine a 'monitoring' and 'diagnostic' function.

The following sections discuss the relevance of gender differences to tests which are designed for purposes of comparison rather than diagnosis.

4.1 The Use of Test Results

It is difficult to predict how the dissemination of test results to parents will affect their actions and perceptions. In terms of local or central government actions, however, there are several clear possibilities. An LEA could, for example, use test scores as indicators of need and direct resources to schools with poorer performances. Alternatively, it might take the view that poor performing schools should become, say, candidates for closure or amalgamation.

Gender differences will be an important factor where there are disparate proportions of girls across schools, and of course in the case of single sex schools. If the tests produce, for example, lower average scores for girls, then those schools with high proportions of girls will tend to produce lower mean test scores. If such schools 'suffer', either through parent or LEA action then it could be said that gender discrimination was occurring, since relatively more girls than boys would 'suffer' as a result. One way to avoid such an outcome would be to carry out a valid statistical adjustment for gender differences when comparing schools. This will be elaborated upon in the next section.

The other relevant aspect of the use of test results is in individual selection and placement. Thus, a secondary school might have ability groupings for its classes, say for mathematics from the second year. If national mathematics tests scores are available on each child, there would be some pressure, presumably, for these to be used in the ability grouping procedures. In some cases schools may find it difficult to carry out ability groupings in any other way when pupils and parents have access to each child's test results. As pointed out in section 2.5, tests can be constructed generally to reflect any desired degree of gender difference (or none at all) with consequent effects on the numbers of girls and boys in different groupings.

It is generally accepted that there are advantages which follow from belonging to a high ability grouping, so the possibility of discrimination again arises. The problems of constructing gender-fair tests are discussed below. The principal issue here is that there are no generally accepted objective external criteria for test construction which can guide a test constructor in choosing the
relative weightings for different test questions or topic areas. If it were argued, on grounds of equity, that tests should therefore be constructed to give equal average scores for boys and girls, then consistency would require that this should be done at all ages. Furthermore, such a procedure would raise questions about other assessment procedures such as the GCSE and whether attempts should be made to balance results by gender more generally in relevant public exams.

It is worth pointing out that this issue cannot be resolved by appealing to historical precedents in test construction. What has occurred previously partly reflects the culturally related expectations of test constructors concerning gender differences (Gould, 1981).

4.2 Adjusting for Gender Differences

If gender differences are eliminated, in the manner suggested above, or by standardising scores according to separate gender norms, then differences between schools would not be expected to depend on the proportions of girls in the schools. If, however, gender differences in test scores are allowed to persist, then a complicated statistical adjustment would be necessary. A discussion of the problems surrounding such procedures is given by Goldstein (1987).

Suffice it to say here that the use of individual children's test results is required rather than school or LEA averages. It should also be noted that there are other factors, such as ethnic background, social class and attainment at time of entry to school, which need to be adjusted for in order to make fair comparisons between schools or LEAs.

It seems that the complexities of the process effectively would rule out its routine use. Certainly it would require a level of expertise unavailable to most LEAs. Moreover, even if such adjustments could be carried out routinely, similar arguments concerning consistency apply here as in the case of test scores used to compare and group pupils.

4.3 Gender-Fair Tests

It is now commonly accepted good practice that test constructors have an obligation to ensure that their instruments contain no sexual or racial stereotypes or any material which could be offensive to a particular group. There exist guidelines (American Psychological Association et al, 1985), and any national system of testing should aim to follow these broadly. It would also be important that the details of test construction are publicly available so that public scrutiny and challenge is possible if guidelines are not followed.

Of course, even where such guidelines are followed, these cannot be expected to eliminate all gender difference, so that their adoption does not remove the necessity for considering the issues of sections 4.1 and 4.2.

5. Further Work

The review of the evidence of section 3 has not attempted to evaluate existing research findings in terms of their quality. Some studies have been excluded on the grounds of unrepresentativeness, and it is clear that a more thorough review and evaluation of findings, especially concerning narrowly defined domains of achievement, would be useful. Such a review should also pay
attention to quantifying differences, so that their relative importance can be assessed.

Some other issues are relevant also. The performance of boys and girls in single and mixed sex schools has been the subject of research, but the existing evidence does not seem to be clear-cut and needs careful evaluation. There have been several studies of teachers' expectations in relation to gender differences and this may well be important if teachers are allowed discretion in the administration and marking of tests. A review of public examination entry policies and results would provide further useful evidence on this issue.

Finally, whatever decisions are made about the implementation of national tests, it is extremely important that there is a proper, and independent, evaluation of their effects with respect to gender as well as other factors. Given the uncertainties about the likely effects, such an evaluation could at least be used to alert people to serious problems and possible violations of equal opportunity legislation.
6. References


**Acknowledgement**

This submission is based on a research review undertaken by Professor Harvey Goldstein on behalf of the Equal Opportunities Commission.
APPENDIX G

Technical note on reliability, validity and error

Reliability refers to the scores assigned to pupils and is measured in terms of their stability over time or over different sets of equivalent tasks or over other variable assessment conditions. Computation of a reliability coefficient allows the estimate of an error of measurement: a statement of the range of fluctuation which is likely to be observed in a single pupil's score as measurement circumstances change.

Changes in the circumstances of measurement will usually have direct educational implications in national assessment. The formative purpose does not require that assessments are stable over long periods of time: quite the contrary, they should be sensitive to changes in the educational experience of the pupil. The definition of the national curriculum in terms of attainment targets will often mean that equivalence over different content areas is an objective of the national curriculum, not a prerequisite. For these reasons, arising directly from the differences between achievement and ability testing, we do not recommend standard errors of measurement as the most suitable basis for the computation of the confidence limits of a score.

Validity relates to an interpretation or use of test scores. It is obviously reduced by errors in the estimate of the scores themselves but is further reduced by unwanted influences on the relationship of the measures to the construct which they are interpreted to measure. We feel that the confidence limits of national assessment scores should reflect these influences which largely arise from the necessity of sampling achievements. Such additional sources of error are readily estimated and taken into account in generalisability coefficients (Cronbach et al, 1972; Brennan, 1983; Johnson, in press).

There are three other advantages of the generalisability approach. National assessment is to serve the needs of different users who will wish to interpret essentially the same database in different ways. Different generalisability coefficients can be derived from the same database according to the intended interpretation. National assessments will produce pupil profile reports. Generalisability deals with the overall degree of confidence in a profile. Generalisability theory can be used to refine the data collection techniques in national assessment by showing how best to minimise threats to confidence with the resources available. It seems entirely appropriate that a formative system should use a formative approach to quality control, which is provided by generalisability theory.

Users will want to be able to make interpretations and draw conclusions with confidence but cannot be expected to be familiar with the translation of reliability or validity indices into confidence limits. We suggest that where appropriate these be reported directly, especially in the translation of standard assessment marks to levels and in evaluative reports on school performance. Of course, the analysis required to produce data for users will have to be undertaken by expert agencies, but they could then advise users on the degree of confidence that they could place in the results.
References

Brennan, R-L. (1983)
*Elements of Generalisability Theory*
The American College Testing Program

*The Dependability of Behavioural Measurements*
New York: Wiley

Johnson, S. (in press)
*Technical report on the APU Science Monitoring Programme*
Department of Education and Science
APPENDIX H

The conduct of group moderation meetings

This account is speculative. It is intended to help readers to understand our intention in recommending group moderation. It is not intended to establish firm rules and procedures – which are for determination at a later stage. It is based on experience of the use of group moderation in public examinations.

1. Organisation

For the administration of national assessments, there will be need for a regional organisation between the SEAC and Nee and the schools which will act as a communication link. As moderation is the major channel of communication that link, referred to hereafter as 'region', will have moderation meetings as a prime focus of its attention. This report makes no recommendation about the body which should be responsible for this regional organisation.

2. A chairman should be appointed by the region to act as a link with moderation groups. It is possible that in the early years these could be full-time appointments with general responsibility for INSET especially at the primary level where there is little experience of formal assessment procedures. We have not designated a national assessment day. At age 7 certainly, and at age 11 probably, all assessment is likely to be distributed over time allowing one chairman to deal with a reasonably large number of groups and to carry standards and insights from group to group.

At other times of the year a full-time appointee can:

a. Visit standardising meetings held within and for individual schools
b. Give direct training in schools in the use of standard assessment tasks.

Wherever there is more than a single teacher involved, schools should be required to hold an internal standardising meeting (see Section X) prior to the submission of internal assessments to the group moderation meeting. All teachers responsible for the assessment of candidates are required to take part in an internal standardisation which has an important part to play in school self-evaluation and development. There will be a simple group meeting for age 7, and another for age 11. At age 14 there will be separate meetings for each subject. Similarly at age 16 there could be meetings to determine results for non GCSE pupils, but it is possible that the results for these could be considered together with those for age 14.

The meeting may be organised at the discretion of the school in order to suit its particular circumstances. Chairmen should notify centres of their intention to attend the internal standardising meeting.

The meeting should include activities which promote discussion about standards and enable the centre to achieve an agreed standard of marking. The purpose of a visit by a chairman is to give teachers opportunity to discuss not only the work presented but all aspects of the syllabus support material or administration that may require clarification. The exercise is designed to help communicate national requirements in terms of activities and standards.
3. The group will appoint a secretary to take care of incoming and outgoing documentation. The success of the system will depend upon the quality of documentation which will comprise:

a. A special report form for each school which allows the display side by side of teachers’ ratings, and of profile component scores derived from the standard assessment tasks. This should be pre-printed every year so that raw scores on standard assessments can be lined up with level scores on teachers’ ratings.

b. A pupil report form showing for each component the teacher’s rating, the standard assessment level and a space for an agreed level report or an indication that follow-up assessment is required (Documents B).

c. A form for reporting comments on the pedagogical, curriculum design and assessment implications of components discussed at the meeting.

Computer support would also be advisable. Transmission of data on disc will lead to rapid collation and dissemination of information. The investment in software will also facilitate the internal record keeping which is increasingly required for the initiation of Records of Achievement. Many schools already have arrangements with public examining groups to register their candidates by electronic means. Some generalisation of these arrangements is required.

4. Prior to the group moderation meeting

Groups of schools are established in such a way that each group forms a reasonably compact and stable geographical unit.

One document, Document A, will be prepared from the documents described in 3(a) above, aggregated by the region for the group and for all groups in the region.

The main aspect of interest is the extent to which levels in individual profile components have proved easier or more difficult to attain on the standard assessment tasks and tests. It can be anticipated that there will be a less than perfect match between teachers’ ratings and the test results and it is that pattern of mismatch which is used in the moderation process to identify schools the assessments of which depart from expectation.

Decisions will be taken at regional level, on the basis of the region’s aggregated results, on which profile components and/or standard assessment tasks require detailed attention at group meetings.

It is likely that at age 7 attention will focus on just the one of three assessment tasks which all in a group will have used in common, in order that the richness of contextual effects should not be swamped by an attempt to cover too many issues. At age 11 two assessment tasks should be covered in this way. There will probably be a three year cycle to ensure systematic curriculum coverage.

The group chairman will liaise with the group secretary to plan a format for the meeting that provides for the full discussion of the standard assessment material and of issues on the agenda likely to arise.

All members will receive personal copies of documents in the form of 3(a) above, for each school in the group, for the group as a whole and for the region as a whole.
5. **Conduct of the meeting**

Group moderation meetings are crucial in encouraging positive views on the syllabus and the assessment procedures. Thus the promotion of positive discussion of the materials, the drawing in of new and young teachers and the highlighting of principles and good practice are likely to facilitate not only the assessment procedures but also the continued success of the scheme. These meetings are the initial forum for the debate of issues and principles.

*First phase*

This will consist of a scrutiny by all present of document A from the group related to document A for the region.

The group will consider first at the regional level the relationship of teachers' ratings to standard assessment results component by component. Where there are sound reasons for a discrepancy between the two these will be recorded on a further document, document C, and a decision taken as to which is the better estimate of the component. Otherwise the standard assessments will stand.

The group will then compare both (a) the group's pattern of teachers' ratings and (b) the standard assessments with the agreed regional outcome. This consideration is of the two patterns of discrepancies across profile components, NOT of the absolute standards. Statistical moderation will highlight profile components requiring special consideration. Again any special reasons for departure from the regional pattern will be recorded. Otherwise it will be expected that the group follow the regional pattern.

Finally the group will consider the group's overall standards vis a vis those of the region. Again, statistical moderation taking into account demographic variables can serve to focus the group's attention and again any justification for departures from the regional standard will be considered.

By the end of this phase the group will be alerted to particular profile components where some adjustment by the individual schools within the group will be necessary to bring the group into line with regional standards.

In this phase discussion of the standard assessment material should provide teachers taking part with the opportunity to raise and discuss any other matters or problems relevant to the assessment. They may wish to seek guidance from the region on certain issues or to make proposals for amendments of amplifications to the conduct of national assessment. Matters raised in this way by groups would normally be discussed in the first instance by either the region or be referred, if necessary, to SEAC. The SEAC may also use the channel of the trial meetings to inform teachers of relevant matters or to request their views on aspects of the scheme of assessment.

*Second phase*

The group now considers each school's pattern of outcomes in relation to the expectations
now established for the group as a whole. Sometimes all schools in a group may have to accept an adjustment to one or other component. That aspect will be emphasised and its assessment discussed in detail. At other times one or two schools may have been identified by statistical means as at wide variance in their marking and the chairman should, as tactfully as possible, invite them to comment on their assessments rather than remain in silent disagreement with the general feeling of the group.

It is essential that all teachers in the group have the opportunity to share their standards and experience of assessment with other teachers in the group. This process not only helps to establish comparability but also encourages staff development in assessment.

Under the guidance of the chairman the group may wish to discuss administrative issues connected with the submission of work and grades that might be helpful, particularly to new centres.

At the end of this phase the group must ratify adjustments school by school. These adjustments are in terms of percentages of pupils at a level. For example, an adjustment decision might require a school starting with 30% in level 1, 50% in level 2 and 20% in level 3 to alter the distribution to 25% in 1, 50% in 2 and 25% in 3. It is for the individual school to determine which pupil assessments have to be changed to bring the school into line with the distribution approved by the group.

Third phase

The major patterns of individual pupil records will be considered and problems arising in relation to exceptional pupils, who will be indicated as such on Document B (see 3(b) above). The group will discuss the circumstances in which follow up assessments are required and notify the secretary of the need for follow up materials. Decisions on the first level assignments for individuals will however be the responsibility of the school.

All agendas will include standard items of business, such as the assessment or teaching or the administration or regulations of the national curriculum. These may arise from the agenda or from a paper such as the annual report of the chairman of chairmen or from items of concern to the group. The group chairman, in planning the format of the meeting, may consider inviting experienced members of the group to lead, or make structured contributions to, such discussions. A key aim in these parts of the meeting should be to promote the sharing of the best practice as well as full understanding of the national curriculum.

After the meeting

Each school submits amended Documents A and B to the chairman for ratification prior to reporting results to parents.

Group chairmen will be invited to an annual meeting at region to discuss issues which have been raised by the groups during the year. It is important that group secretaries record such matters in the minutes of the group so that they can be included in the agenda for the annual chairmen's meeting.
APPENDIX I

Records of achievement

A record of achievement brings together the wide range of a pupil's achievements and records progress and activities both within and outside the classroom, including experiences and achievements reflecting qualities not tested by examinations.

The Government's consultation document: "the National Curriculum 5 – 16" recognises that the national curriculum will not remove the need to place a value on all of a pupil’s achievements and to record and report these succinctly for the benefit of pupils, parents and employers. It states the Government's commitment to the national introduction of records of achievement in the following terms:

"The Government aims to set in place by 1990 national arrangements for the introduction of records of achievement for school leavers. Such records ... will have an important role in recording performance and profiling a pupil's achievements across and beyond the national curriculum (paragraph 32)."

Education Support Grants are being used to support development work in 9 pilot schemes involving 22 local education authorities and some 250 schools in England and Wales. The object of the schemes is to cast light on the best way to set about recording pupils' achievement with a view to establishing the greatest possible degree of agreement on crucial issues and to prepare the way for national guidelines.

Support for the pilot schemes has been provided over the three years 1985-88. A National Steering Committee was set up to oversee and monitor this development work. The schemes will make their final reports in March 1988 and the National Steering Committee will produce a final report in the Autumn of 1988. This will offer the Government draft national guidelines within which records of achievement can be introduced for all young people in secondary schools in England and Wales. This report will be published.

The National Steering Committee produced an interim report in November 1987, copies of which can be obtained from the Department of Education and Science, Publications Despatch Centre, Government Buildings, Honeypot Lane, Stanmore HA7 1AZ.

The evidence submitted by the National Steering Committee to the Task Group on Assessment and Testing is reproduced below.
Records of Achievement and the assessment of the National Curriculum: evidence to TGAT from the National Steering Committee for Records of Achievement

Introduction

1. Our aim in making this submission to the Task Group on Assessment and Testing (TGAT) is to explore the close interrelation between TGAT’s terms of reference and the pilot development work on records of achievement (ROA) which we, the Records of Achievement National Steering Committee (RANSC), are steering with the objective of advising Government on the establishment by the end of the decade of national arrangements for ROA for all young people in secondary schools.

2. The background to our establishment, and the policy framework in which we are operating are set out in "Records of Achievement: A statement of policy" published by the DES and Welsh Office in 1984 and hereafter referred to as PS. We are overseeing ESG-funded pilot work in 22 LEAs in England and Wales and have established links with the other major development initiatives covering most of the remainder. We are due to report in autumn 1988 and will be issuing an interim report later this year.

3. The Government’s continued commitment to this important policy initiative was re-stated in paragraph 32 of the National Curriculum consultation document (CD). In that paragraph, the proposed arrangements were described as "recording performance … and profiling a pupil's achievement across and beyond the national curriculum." The phrase "across and beyond" gives a clear indication that the concerns of ROA partly overlap with those of the National Curriculum (NC) and are partly separate from it. This paper concentrates primarily but not exclusively on the former.

Common aims and objectives

4. Both ROA and NC aim to raise standards. ROA seek to achieve this by giving credit for pupils’ achievements, by increasing pupils' motivation, by identifying their individual curricular needs and by clarifying learning objectives (PS paragraph 11). This is fully compatible with the NC aims and purposes.

5. Though not explicitly mentioned in the PS, several of the pilot schemes have seen the need to specify clear short term objectives in pupils' learning, both to improve motivation and to make what is recorded more meaningful and precise. This practice is in line with the setting out of programmes of study and attainment targets proposed in the CD (paragraph 8 (ii)).

6. ROA share with the NC the assumption that all pupils should have access to a broadly based and largely common curriculum linked to pupils' own experiences. (CD 8 (iii)). ROA are especially concerned to record those experiences which occur during or in connection with compulsory secondary schooling. These experiences were originally envisaged as important to record for their own sake; now, in the light of the NC proposals, they assume even greater importance as a means of providing the context within which assessment against national attainment targets can be set.
7. The ROA initiative is also at one with the NC proposals in stressing the importance of continuously checking on pupils' progress throughout their schooling. ROA see the diagnostic process of stretching pupils further when they are doing well and giving them more help when they are not – CD 8(iv) – as the desirable and likely outcome of the systematic discussion and recording of pupils' achievements during the formative stages of recording (PS 35). These checks on progress – or formative assessments – are of course aimed at supporting the curriculum. We know that TGAT will be seeking to recommend assessment arrangements that similarly underpin, rather than distort, curricular objectives: the evidence from our pilot schemes as to how this can be achieved is of value here, and we will therefore be making available to TGAT our own interim report, together with the interim report from the Joint Open University/Bristol University evaluation of these projects.

8. Though the work of the pilot schemes is largely confined to the secondary phase, there are implications for the primary phase. The PS suggests that ROA should begin with a summary of pupils' experience at the end of primary education. The NC assessments at 11 would provide a valuable base for this summary, but evidence from the secondary phase suggests that these assessments would benefit from being set in the context of a broader statement of achievement. In this context, TGAT may wish to note that many teachers in our schemes would argue that the ROA formative process is one that should be begun before pupils reach secondary schools: even the age of 11 is comparatively late to introduce such a major change in approach to learning. The NC proposals, recommending continuity from 5 – 16 and accompanying assessment in both primary and secondary phases, accord well with these views. We would encourage TGAT to consider the relevance of the ROA process to the primary phase and perhaps to recommend that some appropriate pilot work be undertaken.

Possible tensions between ROA and the NC

9. In several important respects the emphasis of ROA differs from that of the NC. In the PS and confirmed by experience in schools "it is essential that pupils be closely involved in the recording process" (PS 36). This involvement of pupils in their own assessment by means of dialogue between pupils and teachers is a key feature of ROA but a concept that is not explicitly mentioned in the CD. We recommend to TGAT this key principle of pupil involvement which can both inform teachers' assessments under the national curriculum and also play a vital role in giving due emphasis to the diagnostic purposes of the assessments envisaged by the NC (CD 28).

10. Secondly, the underlying principle of ROAs is that, though both the strengths and weaknesses of pupils will be discussed during the formative process, the content of the summary document will be expressed in terms of positive achievement. If ROAs are to be the vehicle for recording and reporting NC assessments, it would be important for the results of these assessments to be in terms of varying degrees of positive achievement rather than to be of a pass/fail nature.

11. Thirdly, the NC proposals, whilst aiming to raise the standards of individual pupils, are also concerned with the accountability of schools and local education authorities. The ROA proposals are essentially for the benefit of individual pupils. This raises two possible
tensions. First, certain pilot developments are strongly committed to the view that pupils are the "owners" of their ROA, having discretion about what to include and to whom to show them. The use of the summary document at age 16 – and perhaps an interim summary document at age 14 (see paragraph 17 below) to convey the results of assessments to parents would conflict with this philosophy. However most, including RANSC, would incline to the view that this principle need only be inviolate in respect to the content of the pupil's personal statement – one element of ROA. Secondly the recording and reporting of NC assessments in ways that fulfil the proposed evaluative and comparative purposes may sit uneasily within a pupil's individual record of his or her unique positive achievement, without reference to other pupils' achievements.

Finally, there is the question of what validation of ROA is required and/or what accreditation of schools. RANSC has been considering this difficult issue, in particular the extent to which assessment of pupils' educational attainment other than by public exam, for instance through the use of unit accreditation or staged assessment, will require external validation. In the light of the NC proposals for the moderation of assessments against attainment targets, a simpler form of accreditation/validation for ROA can now be envisaged, concerned primarily with the overall structure and processes of ROA and the personal and social elements. It will be important that the eventual arrangements for moderation of NC assessments and validation/accreditation of ROA are complementary rather than duplicating.

Issues of Assessment

13. RANSC, like TGAT, is not primarily concerned with the content, articulation or delivery of the curriculum. Both are however concerned with its assessment and have a common interest in the resolution of a number of issues.

14. A particular concern of some pilot ROA schemes has been the assessment of cross-curricular skills and achievement. Relatively modest progress has been made to date in this field, but the issue has been given a high priority because these skills are seen by teachers as a key aspect of student success, both across the subjects of the curriculum and subsequently in the world of work. Problems of definition, and of organisation in schools which tend to be divided into single subject departments, have hampered development of what is for most anyway a challenging and unfamiliar task, and many schools have yet to develop organisational structures which can facilitate the consideration of cross-curricular issues. But ROA have the exciting potential – not least in this matter of cross-curricular skills – to give due emphasis to pupils' total achievements across the curriculum as a counterweight to any bias towards over-emphasis on performance in single subjects. Furthermore, the consultations among the staff concerned with these general learning skills – a pre-requisite, of course, for their assessment – are a most valuable step towards securing coherence between subject departments as the school's curriculum is developed.

15. Projects have been tackling in various ways the assessment of a set of objectives within subjects at various points during pupils' learning. Some projects use a unit credit approach with regular points of assessment; others are involved with staged assessment where, at least in principle, pupils are assessed when ready. The experience of schools involved in these initiatives will be of particular value and interest in relation to the NC concept of assessment and testing at predetermined ages, and we are glad that TGAT has already
held discussions with some of those concerned. Equally the NC attainment targets and their associated assessments should give schools clear guidance on approaches to assessment and obviate the need for every school to devise its assessment policy from first principles. Nevertheless it will be incumbent on schools to devise a coherent policy which embraces ROA, NC and GCSE assessment in a consistent approach. To assist them, more work still needs to be done to bring together the basis for awarding GCSE grades with criteria used for internal assessments and the NC attainment targets.

**Issues of recording and reporting**

16. PS 38 draws attention to the need for schools to bring together with ROA as far as possible the various kinds of reports sent to parents, employers and career officers. Schools have certainly found this to be necessary in practice if the time needed for recording and reporting is to be kept to manageable proportions. We believe that this requirement should extend to the recording and reporting of NC assessments, not only for the sake of the workload in schools but also because the proliferation of different kinds of reports for parents and others could hardly be helpful.

17. The growing realisation by schools involved in ROA that they need a unified and coherent policy of assessment and recording has led to the emergence of the concept of the "interim summary" sent to parents in place of the traditional school report. The incorporation of assessment at age 14 into the appropriate interim summary should pose no difficulties other than that of avoiding negative response to which reference has already been made.

**The wider purposes of ROA**

18. As stated at the start, this paper has concentrated primarily on areas of mutual concern to RANSC and TGAT. It is important however to stress that in many ways the most innovative aspect of ROA is the inclusion of personal and social achievement and experience within the framework of reporting. In this very important way the aims of ROA are different from but complementary to the aims of the NC. The kinds of assessment being developed for personal achievements are very different from the nationally prescribed tests or moderated assessments envisaged as necessary for most of the national curriculum.

19. ROA, by giving weight to those aspects of the formal, informal, or extra curriculum which lie beyond the foundation subjects, can present pupils’ NC achievements within the context of the totality of their achievements and experiences in and outside the classroom. They can therefore redress any tendency on the part of those who make use of the NC assessments to ascribe undue importance to achievement against predetermined subject attainment targets.

20. ROA can provide attestation of those pupils’ achievements that are unforeseen but worthwhile, unplanned but fruitful. There is growing evidence that employers appreciate the importance of more general, well founded statements about young people than can be deduced from aggregation of performance in a number of school subjects.
Conclusion

21. The issues as complex as those relating to ROA and the NC cannot be comprehensively explored in a short paper. We would therefore recommend continuing liaison between TGAT and RANSC to further the debate and clarify the issues. We believe that this would be of benefit to both bodies.

22. At this stage, we will conclude by re-stating the most important point that RANSC wishes to register. We see it as vital that ROA and the NC develop in complementary and mutually supporting ways. The key contributions of ROA to the NC are:

   a. The formative process, with pupil teacher dialogue essential to secure the Government's objective of diagnostic NC assessments; and

   b. The setting of those NC assessments into the context of the totality of a pupil's achievements and experiences.

October 1987
APPENDIX J

Technical note on background factors affecting educational performance

1. A strong association between social background and educational performance of almost all types is one of the longest established and best supported findings in social and educational research. Why this should be so and how pupils' social background may be measured and its effects compared remain much in dispute. Nevertheless it is now widely accepted that social background factors must be taken into account in any attempts to evaluate educational performance and effectiveness whether of individual institutions, or of LEAs. Without such analysis the effects of variations in social background are liable to conceal the effects of school, teacher or other educational variables.

2. Recent examples in this field include work by the Statistics Branch at the DES, which in a series of publications (Statistical Bulletins 16/83 and 13/84), has estimated the contribution made by various factors to variations in secondary examination performance by LEAs.

3. The study reported in SB 16/83 concluded that the two socio-economic group (SEG) variables (representing the social composition of the resident populations of whole LEA areas) accounted for the largest amount of variation between LEAs' examination results, ranging for different measures from 3/4 to 2/3 of the whole variation. In this study these two factors outweighed the effects of other social factors as poor housing, one parent families, unemployment, large families and free school meals. Relative levels of expenditure were even less influential in "explaining" (in the statistical rather than the educational sense) variation.

4. The most recent study confirmed that social background factors – in this case measured by the proportions of the population in high or low socio-economic groups from the 1981 Census – were easily the most important, "explaining" as much as 75% of the variation in LEAs' examination results. There was a strong association between the proportion of the population in high SEGs and the proportion in higher levels of performance (A level and 5 or more higher grades in O level/CSE), and low SEGs and lower levels, particularly the proportion of leavers with no graded result. In comparison, and in line with similar studies elsewhere, school or educational variables such as expenditure or pupil-teacher ratios "explained" very little of the variation in the LEA results.

5. The studies quoted above led to the conclusion that "any comparison between the average attainment of pupils attending secondary schools in different LEAs would be seriously misleading if it failed to take into account the social background of the pupils". Each concludes with a table comparing the actual and "fitted" or expected values, the latter representing an adjustment on the basis of the relationships between examination attainment and socio-economic factors.

6. It is possible to make these adjustments for all LEAs, because of the existence of comparable data for all areas from sources such as the national census. Similar data does not yet exist at the school level across the country. However, the DES has conducted a
pilot study of the effects of social background at the level of the individual school (DES: *Pilot study of school examination performance and associated factors*, 1984), and some LEAs have been able to use school level data to take account of the effects of variations in intake on examination performance at secondary level. The ILEA, for example, using a similar procedure to that employed by the DES at LEA level, has presented results for individual secondary schools which take some account of variations in intake (ILEA: *Looking at School Performance: Actual and Predicted Examination Scores*, 1986). This analysis included data on pupil ability at intake to the secondary school, which is an important component in assessing school performance.

7. These studies at the individual school level demonstrate how data at the LEA level may tend to reduce variations by aggregating results from successful and unsuccessful schools alike. Finer grain estimates of school performance, principally based on examination results – which are not the only criteria valued by parents, employers and teachers – show considerable variations between schools, when social background and other intake differences such as ability and attainment have been taken into account. A series of research studies since the pioneering work of Rutter (*Fifteen Thousand Hours*, 1979) has confirmed that significant differences in the effectiveness of schools emerge once the data is examined at this detailed level (or "disaggregated" in research terms). Recent studies, for example the ILEA Junior School Study (ILEA, 1986), have suggested that individual pupil "progress" – that is change in attainment over time – may be more closely correlated with educational factors than with social background. However, social background remains the most important factor in explaining variations in performance at the group (for example school or LEA) level at anyone point in time.
APPENDIX K

Numbers in primary schools

1. Statistics for England and Wales on Four Year Olds in Schools

The latest figures available for England (DES Statistical Bulletin 9/87) show that in 1986 there was a population of 1,175,000 children aged three and four. Of these 235,867 (20 per cent) were attending infant classes and 246,893 (21 per cent) were attending nursery schools or classes. (There were also 25,587 two–year-olds in nursery provision.)

These figures represent an under-estimate of the number of four-year-olds attending school, as they exclude 'rising fives' (i.e. those aged from four years nine months to four years eleven months). There was a slight decrease in the number of four year olds in schools in 1986 when compared with data for the previous year, but this is accounted for by a drop in the four-year-old population in 1986.

Figures supplied by the Welsh Office show that, in January 1987, there was a population of 70,600 children aged three and four. Of these 15,886 (22 per cent) were in nursery classes compared with 33,234 (47 per cent) in infant classes. (There were also 2,497 two-year-olds receiving nursery education.) There has been a tradition of admission to school before statutory age in many Welsh authorities which accounts for the high proportion of the three and four–year-old population in infant classes.

Other useful sources on this are:

a. Results of an NFER research study Starting School at Four conducted in 1986.

b. NFER have just published the report of a joint NFER/ISCDC seminar held earlier this year: NFER (1987), Four Year Olds in School: Policy and Practice (NFER: Slough).

c. A new NFER project, aiming to research the curriculum needs of four–year-olds in school, will start in January 1988.

d. An NFER conference will be held on February 18th 1988 entitled Starting School at Four: Planning for the Future.

Reference

Department of Education and Science (1987)
Pupils Under Five Years in Each Local Education Authority in England – January 1986
Statistical Bulletin No. 9/87
APPENDIX L

How the system might work in a primary school

1. This appendix elaborates all the features summarised at the end of section XIII on the primary stage. It incorporates these into a more detailed account in order to convey the possible implications and consequences of our recommendations for the work of primary schools.

2. At ages 7 and 11 the basis of the reported assessments would be the combination of informal, routine assessment carried on in the classroom (enhanced by clear attainment targets and programmes of study), and of the formal standardised tasks (tests) to be set towards the end of the year. Thus the standardised tasks would constitute one part of the assessment instruments used in the primary phase. It would be possible to support the informal assessment with suitable tests set at the school's discretion; these might include diagnostic tests where the head teacher and class teachers deemed it helpful to confirm, amplify or amend judgements previously arrived at either in the informal assessment or at the formal stages of reporting. A fuller picture of a child's progress would derive from work done and assessed between ages and reported at more frequent intervals to parents.

3. We do not expect, however, that this assessment system will be additional to the many others that now exist, often side by side in the same school. We hope that the need for these – and especially for norm-referenced tests which bear little relationship to the curriculum – will disappear with the introduction of a flexible assessment system related to curriculum and progress.

Age 7 tasks

4. Towards the end of the year in which pupils become 7 the teacher would be asked to select three standardised tasks using a list of titles provided to describe topics available in a nationally prepared set of such tests. The range of tasks in the bank would enable a selection of topics to be made which would be within the children's experience, background and culture and which could be fitted unobtrusively into the run of the work being done at the time.

5. Each of these tasks would assess profile components covering knowledge, skills and understanding in numeracy, literacy, science, physical skills as well as other general skills such as observation or art. Language and mathematics, for example, might each have several profile components appropriate to age 7.

Some profile components would be identified with subjects; others would represent more general competences. Assessment for these would allow for a variety of forms of presentation, and pupils might produce responses for assessment by written work, or by oral work, or by actions which the teacher observed in a systematic way. Thus the tasks would be in keeping with primary practice, taking account of the nature and stage of young children's learning and their immediate experience while offering for the teacher clearly delineated goals and standard methods of assessing the performance produced. Pupils
would be able to tackle the tasks at a variety of levels so that differentiation would be by outcome. Some tasks, or parts of tasks, might be directed at individual pupils, working on their own, others for pairs or groups. Examples of two such tasks are given in Appendix D.

6. At least one of the three tasks should be a choice agreed by a group of schools so that moderation can be undertaken on common ground. Later, when teachers have benefited from in-service training and are experienced in moderation, all three tasks might be at their choice so that they might select tasks which best suit their pupils’ circumstances.

**Grading**

7. We expect that the work of the great majority of pupils at 7 would fall within three levels and performance would be recorded at levels 1, 2 or 3, though it would be possible for exceptional performance to rank higher than this. We envisage that most children would achieve level 2 at the age of 7. Children attaining level 1 would be those identified as needing extra help because they have made little progress. Those achieving level 3 would be children in need of work of greater pace and depth in order to use their abilities to the full. In both these cases the school would be able to make use of appropriate diagnostic tests where it considered them suitable.

**Moderation**

8. A teacher who has collected results for all pupils on the three tasks would be able to assign back to level 1, 2, or 3 on the basis of the results for each profile component; from these the percentage of the total who are at each of levels 1, 2, 3 could be derived. Such results could be compared with the corresponding results for the teacher's own classroom ratings, accumulated over the year. The teacher would then go to a moderation meeting where a group would exchange and compare such results and compare samples of work, in order to decide whether and how to scale the results. At the end of this, there would be an agreed overall distribution of pupils across the three levels. Subsequently the teacher could reconsider the level assignments of each of her or his own individual pupils and decide thereby on the report to the parents of each pupil.

**Age 11**

9. At age 11 a similar strategy would apply. Performance would be assessed using a combination of informal assessment carried out over the four years but for reporting purposes concentrating on the last year, and the formal standardised tasks set on a national scale.

10. We envisage that there would be three or four standardised tasks, again selected by the teacher from an item bank to suit the run of the work. The main difference at this stage is that the tasks would be more complex and would need to take account of the larger number of profile components which would become increasingly subject related. Although in the main differentiation would again be by outcome, it would be possible to draw on some standardised tasks offering different demands within a given profile component. There might also be more narrowly focused tests to yield information on particular profile
components and so supplement the data from the task tests.

11. At 11 most pupils’ performance across the profile components would fall within the levels 3, 4 and 5. All these represent progress since age 7 and we expect that most children would achieve level 5. Those obtaining level 5 (or higher) at this age would be deemed to need extra support if they are to continue to maintain their progress. Those obtaining level 4 and below would be identified as in need of help. The use of the profile components would enable teachers to identify to particular strengths and weaknesses. Any pupils who have not advanced by at least one level since age 7 would clearly need special attention.

12. As for age 7, results for each profile component would be assembled both from the national tests or tasks and from teachers' own ratings. The overall positions of pupils at each of the levels would be calculated. Teachers would then go to a group moderation meeting to discuss their results and consider action on discrepancies. The outcome would be agreed overall distributions, and an opportunity for each teacher to reconsider decisions about individual pupils before reporting to them and their parents.

Special needs

13. At both ages any child who was clearly unable to manage a standardised task (test) unaided could be helped; this help would be recorded and might signify that he or she had not attained the competences required in that profile component. If a "non-statemented" pupil was judged by the head teacher and the class teacher to be quite unable to undertake a given task for medical or emotional reasons, the child could be exempted from it and this exemption would be recorded. At both 7 and 11 children who have been "statemented" would undertake the standardised tasks to the extent specified in the statement.

Reporting arrangements

14. Our proposals for the reporting and publication of aggregated results for classes and schools are described in detail in sections XII and XIII.
APPENDIX M

How the system might work in a secondary school

1. This appendix elaborates of the features summarised at the end of Section XIV on the secondary stage. It incorporates these into a more detailed account in order to convey the possible implications and consequences of our recommendations for the work of secondary schools.

Entry to secondary school

2. Under the national assessment system, all secondary schools would be receiving criterion-referenced and calibrated results based on the collection of data for every pupil on transfer from their feeder schools. It should be clear, from this transfer assessment information, what pupils know, understand and can do after the primary phase of education; and this will be directly related to the attainment targets set for core and other foundation subjects. This information should form a sound basis for whole school and departmental decisions about the learning needs of the new intake of pupils, so that continuity and progression are more likely to be achieved. At least some of the transfer information would be specific to subjects and profile components, so that relevant departments would be able to allocate tasks at the appropriate level of challenge for pupils from the outset of the secondary phase. Teachers would also have indications of the likely range of pupils' future performance to build on at the transfer stage, based on the primary school's assessment of the general knowledge, understanding and skills which have applicability across the whole curriculum at secondary schools.

Course planning

3. Overall course planning and directly-related pupil activities for class and homework, would need to ensure that all the specified subject profile components were covered in a balanced and relevant way throughout the programme of study. Although the system allows for assessment of single profile components to illustrate varying pupil strengths and weaknesses, the overall characteristics and cross-curricular relevance of subjects would need due attention in teaching and assessment practice too. Good practice should also ensure that all teaching groups were properly informed about the relevance and meaning of subject profile components and their assessment, so that pupils were encouraged to be in a positive partnership at classroom level aimed at clearly understood and shared learning and assessment objectives.

Ages 11 to 14

4. By the age of 14, as earlier, the assessment and its recording would derive from a combination of routine teacher assessment of normal class work and homework against clear objectives, with inter-class and inter-school group moderation of selected work samples, and the use of external tests. All this information would determine the levels of achievement reached by each pupil in the profile components of the foundation subjects. The teacher assessment evidence will be built up and refined over the three years from 11 to 14, though concentrated for reporting purposes in the last of these years. The national
tests would be used near the end of the age 14 year. The existence of subject departments would allow for sharing of both the formal and informal work of assessing and recording. Through departmental meetings the assessments, with samples of pupils' work, could be moderated across classes and an overview of the levels reached by each pupil could be built, in terms of the level derived from the teacher's own assessment and the corresponding level results from the national tests. As for age 11, these results would then be prepared and taken to an inter-school moderation meeting. Each such meeting would now be concerned with one of the foundation subjects. The aims and outcomes would otherwise be similar to those for age 11.

5. We expect that at age 14 most pupils who have made progress since 11 would achieve levels in the range 5 to 7, though the span of performance at this stage could be reflected in the award of levels outside that range. The levels for national assessment would be related to GCSE in the sense that achievement within their normal range at age 14 should indicate that, given continued progress, the pupil was on course for grades at GCSE. The review of progress might be aided by reference to each pupil's results at age 11.

6. In the intervening years between 11 and 14, schools might wish to use not only their own curriculum-related tests (whether of their own devising or those associated with the particular courses), but also diagnostic tests – whether nationally provided or commercially available – for cases where specific needs have to be identified. This process should be assisted by the possibility of comparing with results, at age 11, both for individual pupils and for the school as a whole.

7. Levels of progression defined within each profile component should enable teaching practice to aim for appropriate differentiation of tasks and outcome expectations directly related to the pupil stage of achievements, so that sufficiently challenging tasks are set for progression. The shorter-term goals in stages of achievement should also help to promote greater self-awareness of progress and encourage pupil motivation.

8. The use of records of achievement as a vehicle for recording results would mean that the information was set out in the context of general development, giving credit beyond the individual profile components.

Age 16

9. At 16 the great majority of pupils will be entered for GCSE and other nationally approved examinations which are available for certification and qualification purposes. These examinations would however need to be supplemented by national assessment for those pupils not entered in particular subjects, and especially for those who leave school at Easter. As at age 14 the basis for the assessment would be the routine assessment of pupils' work, building on that at 14 and supported by internal and external moderation; standardised tasks (tests) would also form part of the assessment at this stage. The levels would, as at earlier ages, relate to profile components; but for the reasons explained in sections XI and XIV of the report we do not envisage at this stage a precise link with to GCSE grades. The results would be incorporated in each pupils Record of Achievement and the national assessment procedures should thus help to maintain it even where a pupil
Reporting

10. At ages 14 and 16 the results of the assessments would be made available in detail to pupils and their parents, and performance would be set against aggregated national results. At 16 the reporting would be largely in terms of GCSE and other examinations. Publication of aggregated results is dealt with in detail in section XII.

Effects on teaching and learning

11. We expect that methods of national assessment will, with the national curriculum changes, have beneficial effects on styles of teaching and learning. The requirement that individual pupil progress be recognised and encouraged will have teaching implications across departments, calling for a more flexible approach to the processes of learning. There is likely to be an increasing need to develop active learning utilising a range of methods beyond the didactic and into discovery, experiential and problem solving approaches which allow for and encourage individual pupils' progress and state of readiness. In particular, teaching practice will need to give due weight to the levels of achievement which pupils have reached, through the selection of tasks set by the teacher, for example by use of open ended investigation to complement the more structured task with teacher guidance. In practical situations, including but not confined to science, adequate non-teaching support and proper use of technician staff in preparatory work and lesson organisation will be increasingly important so that all levels of ability and progress are properly catered for in the relevant subjects.
Resource needs – some data on existing and planned provision

These data are presented here as a general background. No attempt is being made for this report to draw conclusions directly bearing on the costs of a new national assessment scheme.

1. LEA Training Grants Scheme

In the 1988/89 arrangements for LEA Training Grants Scheme, provision is made for grant-related expenditure of £15m on training in the assessment of achievement. Training which is eligible for grant aid under this heading comprises any training for primary, secondary and FE teachers which is concerned mainly with the development of skills required to "assess, record and report achievement". It is clear from the commentary in Circular 9/87 that most of this expenditure is aimed at secondary work and in particular is geared to development of these skills in the GCSE context. It is worth noting that this £15m is part of a total of £77m identified under various national priority headings (on which grant is payable at a rate of 70%). In addition the 1988/89 programme provides for grant-related expenditure of up to £130m on locally assessed needs (on which grant is payable at a rate of 50%).

2. Education Support Grants: Records of Achievement

Some developmental work is being undertaken with ESG funding through the Records of Achievement project. In England eight development schemes involving 14 LEAs are currently being supported at a cost of £2.25 million per annum. The grants were introduced in 1985/86 and will cease at the end of the 1987/88 financial year. An extension of grant for a further two years will be offered for 5-8 pilot schemes, involving expenditure of not more than £1.2 million in 1988/89 and a similar amount in 1989/90. In addition a further pilot scheme is running in Wales and covers all eight Welsh LEAs.

3. Teachers' Conditions of Employment

The new Conditions of Employment for Primary and Secondary Teachers include as part of the professional duties which a teacher may be required to perform, the assessing, recording and reporting on the development, progress and attainment of pupils. They also include "participating in arrangements for preparing pupils for public examination and in assessing pupils for the purposes of such examinations; recording and reporting such assessments; and participating in arrangements for pupils' presentation for and supervision during such examinations". The conditions cover the number of days (195) and the number of hours (1265) in a year for which a teacher may be required to work at specific times in a prescribed place.

As far as the 195 days are concerned, five of these are days on which pupils are not present and which may be used for, among other things, INSET activities. Some schools/authorities may be using most of these five days for INSET: the general experience seems to be that two or three of them are available for this purpose. In many cases these additional days would only become available at the beginning or end of term and there could well be logistical problems in using these days for a major INSET exercise at LEA level, given the large numbers of teachers involved.
4. **The cost of a day's INSET**

The cost of providing a day's INSET for a teacher will vary from authority to authority depending upon local policies relating to the reimbursement of travel and subsistence expenditure incurred by teachers in attending courses, and upon the arrangements whereby authorities deliver their INSET programmes. The full cost could be as high as £75 per day, on average might be nearer a figure of £50 and in some cases could obviously be below this figure. The size of the teacher force nationally is of the order of 170,000 in primary schools and 235,000 in secondary schools (DES 1985 statistics). On an average basis it would require something over £20m to provide one day's INSET for every teacher where this involved absence from teaching and the provision of supply cover.

5. **External Bodies**

At present, substantial expenditure is incurred on 16+ examinations through GCSE Boards and certain other examining bodies. Most of this expenditure is derived from candidates’ fees for which LEAs are mainly responsible.

There are also two other bodies (the APU and NFER) which are involved in assessment/testing activity. The APU currently operates on a budget of about £1m annually and the NFER on one of about £2.5m.

6. **Education Reform Bill – Financial Memorandum**

The Explanatory and Financial Memorandum to the Education Reform Bill contains the following paragraph:

"The implementation of the National Curriculum will lead to estimated additional public expenditure rising to £33 million in 1990-91. This covers the establishment and activities of the National Curriculum Council and the School Examinations and Assessment Council (clause 7), an element for research and development particularly in the areas of assessment and testing, and the administration of assessment arrangements, for which local authorities will not be charged fees. As the National Curriculum is progressively introduced, some redirection of recurrent and capital resources will be required by local authorities, school governing bodies and headteachers as well as by central government in its planned development of education support and training grants."

7. **Other Data**

Relevant data which might be considered with the above are the costs of development work of the examining groups for GCSE, and the resource used by local authorities, aided in many cases by grants from central government, to establish the graded assessment and records of achievement systems. It may be helpful to consider these, with the above, when consideration of the system needed to support national assessment, is further advanced.