



Qualifications and
Curriculum Authority

Evaluation of the trial assessment arrangements for key stage 1

Report to QCA

The University of Leeds
Assessment and Evaluation Unit
September 2004

The Research Team

Report authors

Diane Shorrocks-Taylor

Bronwen Swinnerton

Hannah Ensaff

Melanie Hargreaves

Matt Homer

Godfrey Pell

Peter Pool

John Threlfall

Other members of the research team

Roger Beard

Veronica Sawyer

John Willcocks

Project administration

Louise Greaves

The team would like to thank all those who gladly gave their time to take part in the trial: teachers, headteachers, LEA personnel, parents and governors.

We would also like to thank the staff at the DfES for supplying the 2003 and 2004 trial schools data used in this report.

Evaluation of the Trial Assessment Arrangements at Key Stage 1

Executive Summary

The project

The DfES document 'Excellence and Enjoyment' (DfES, 2003) set out a proposal to trial an approach to assessment at the end of key stage 1 that uses national tests to underpin Teacher Assessment, with only the Teacher Assessment level being reported to parents, and with LEAs moderating the assessment.

Local Education Authorities were invited to volunteer for the trial, and of those who applied, 34 LEAs were selected to be representative of the national profile. This is approximately a quarter of all the LEAs in England.

The University of Leeds was awarded the contract to evaluate this trial and report the findings to the QCA by September 2004. The purpose of this evaluation is to offer sufficient information to determine the viability and desirability of national roll-out of the assessment arrangements of the trial. The aspects of most importance to this are:

- the robustness of the assessments made - to report on whether the assessments offer data that are as robust as those arising from existing arrangements, especially in relation to their role in value-added comparisons and target setting;
- the attitudes of stakeholders to the new arrangements in the light of the trial - to report on whether the assessment system is felt to be worthwhile by teachers, managers, parents and school governors;
- the resource implications of the trial - to report on what seems to contribute to success.

Methodology and sample

The data collected to evaluate the trial assessment arrangements were of two kinds:

- pupil data from schools, comprising both Teacher Assessment levels awarded and task and test results for Year 2 children in both 2003 and 2004;
- information from and views of stakeholders, using questionnaires, interviews and focus groups, and also observations in schools and LEAs, and materials provided by LEAs.

Much of the data were collected from a sample of 107 schools in 10 LEAs, which were selected and have been shown to be representative of the schools and LEAs taking part in the trial.

- Data regarding task/test and Teacher Assessment were obtained from the DfES for 2003 and from both the DfES and the evaluation schools for 2004.
- Questionnaires were sent to headteachers, teachers, LEA personnel, governors, junior schools, and parents.
- Interviews took place with teachers (22) and LEA personnel (9).
- Focus groups were held with teachers (10) and with parents (6).

Main findings

Robustness

‘Robust’ Teacher Assessment in the trial system, in the relevant subjects, will arise from teacher judgements that have been soundly grounded in evidence, appropriately recorded and reported, and systematically moderated in line with national standards.

- The data in 2003 and 2004 are largely comparable, and what differences do occur (even where statistically significant) are relatively small, and do not exceed the expected cohort differences from year to year.
- In 2004 as compared to 2003 there is a tendency in all subjects for more children to be reported as working at level 2, with correspondingly fewer at level 1 and/or 3. The differences apply when comparing Teacher Assessment in 2004 with that in 2003, and also when comparing Teacher Assessment in 2004 with test and task results in 2003. However, there is also a corresponding change in the results of the tests and tasks for 2004 compared to 2003, with more children awarded level 2 and fewer in level 1 and /or level 3. The patterns in the relationships between results of various kinds are otherwise comparable between 2004 and 2003.
- There is no evidence of any differences in the assessment results from schools of differing types arising from the trial arrangements. Allowing for differences in the samples, Infant/First schools and Primary schools have similar patterns of results.
- Teachers, headteachers and LEA assessment coordinators all felt that the Teacher Assessments in the trial arrangements were more accurate than the reported assessments under the ‘old’ arrangements.
- The confidence of some teachers in making Teacher Assessment based on evidence was low at the beginning of the trial. However, confidence grew throughout the trial as the teachers came to understand and implement the new arrangements.
- Teacher confidence was most easily established in the schools that had well-developed assessment policies and systems.
- Accuracy and teacher confidence in making Teacher Assessments is strongly affected by the quality of training and moderation.
- The evaluation has not yielded clear-cut examples of practices in moderation that are to be avoided. The practices that seemed more effective included: early meetings of moderators to arrive at a consensus about levels and procedures; an emphasis on negotiation rather than imposition (support rather than policing, moderation rather than audit); and clarity among all participants about the range and quantity of evidence needed for moderation.
- The trial appears to have been conducted with good quality training, effective moderation and conscientious teachers, and the performance data from the trial were comparable in most respects with that from the ‘old’ system as operated in trial schools in 2003. As such this suggests that the new arrangements offer a system of assessment that is at least as robust as the system in which tasks and tests and Teacher Assessment are reported separately.

Attitudes of stakeholders

- The majority of stakeholders were positive about the trial, and there is strong support for roll-out among teachers, headteachers, and LEA assessment coordinators.
- Views are varied both across and within LEAs, but there are no parts of the country where negative views are more than a small minority.
- Assessment at key stage 1 represents a heavy burden for schools, but the new arrangements have not been seen as adding to that burden significantly. It seems that workload is not seen as an issue when the work done is felt to be helpful towards an end that is desirable.
- Most teachers have reacted positively to the new arrangements. Many have recognised opportunities to reduce workload, and to enhance their professionalism.
- Some teachers have done more work in the trial than was strictly necessary, though in many cases this was said to have been as a result of insecurity, and the teachers were seeking reassurance by ‘making sure’. Most of these teachers expect workload reduction in subsequent years.
- LEA assessment coordinators have tended to see the new arrangements as an opportunity for professional development of teachers and schools, and any increases in workload for them have been seen as the price of good investment for the future.
- All the respondents to questionnaires felt that the stress and pressure on the children had diminished or had stayed the same.
- Parents tend to prefer their children not to be tested, in order to keep the stress and pressure on them to a minimum, but still like to receive test performance information about schools, including their own child’s school.
- Parents seem by and large happy with the trial arrangements, and there is strong evidence that they have faith in the ability of their own child’s teacher to assess accurately. Many parents nevertheless seem to have less trust in the assessment judgements of teachers in other schools.

Issues in implementation

- Tests and tasks are important as a standard for national calibration, but there are always uncertainties about test and task results in individual cases. Equally, teachers’ professional judgements usually provide the most valid insights into children’s attainments, but there can be bias, relativism and ‘halo’ effects. This points to the importance of checks and balances in the system.
- Given that a very large number of individual teachers are making assessment judgements, the reliability of the assessments will be directly linked to the quality and effectiveness of the training and moderation that is put in place. This needs to ensure the same criteria and standards are being applied across all teachers, schools and LEAs. It is not clear that such thorough and pervasive systems exist at present, especially from LEA to LEA.
- The experience of the trial suggests that the confidence of teachers in making Teacher Assessments grows with experience, and most managers feel that confidence will continue to grow, although a transition period may need to be allowed for, especially where schools do not have well developed assessment policies and practices.
- Some schools already have a successful and supportive ‘assessment community’, and these schools may be expected to be best placed to implement the new arrangements with ease. However, assessment practices rarely operate across groups of schools, and even more rarely

Evaluation of the trial assessment arrangements for key stage 1

across LEA boundaries. In order for Teacher Assessment in its new role to be reliable and robust, it seems desirable to promote cross-school and cross-LEA activities.

- The findings of the evaluation suggest that there are benefits to different kinds of training being offered to different groups. The training offered to 'lone' Year 2 teachers (who may need additional guidance and support to compensate for the absence of other Year 2 teachers), new Year 2 teachers (who do not benefit from contrasts with the past), groups of Year 2 teachers in larger schools (who can benefit more from shared experiences and in-school work), and school managers (who need a greater emphasis on strategic issues) will be more effective if differentiated.
- Training needs to be delivered early in the school year and needs to have clarity in the messages it sends. The 'levelling' of material brought or sent in by schools can be helpful to established shared understandings among teachers. This is particularly important among the teachers who will be used as moderators within the LEA.
- The attitudes of teachers to the new arrangements seem to be improved most by teachers taking part in moderation activities, as only then are anxieties about what is really intended assuaged.
- The varying circumstances in LEAs must clearly be taken into account in the approach to moderation and training that is adopted, so some flexibility of approach will be needed. However, this will need to be much more limited than the complete freedom exercised in the trial, in the interests of both validity and reliability.
- There will need to be some clear thinking on the part of LEAs about how their staff can be helped to move from the 'auditing' role with schools to the 'moderation' role. In some cases this will represent a larger step than in others.
- The LEA role in training and moderation and in supporting schools in their implementation of the new arrangements, which is vital to its success, represents an increase over what is required to successfully audit the existing statutory system.
- There were data collection problems in the trial, but these were in the main the result of changes to software rather than specific trial related issues.
- The management responsibility for ensuring robustness in the teachers' assessments may be felt more deeply in roll-out than it was in the trial.
- School managers need to know early in the year about what is involved, in order to organise the school resources and timetable appropriately. Greater flexibility does not mean total freedom, however, and school managers need clear guidance about where teachers are free to choose, and where they are not, in order to plan accordingly.

Contents

1. Introduction: background and design of the evaluation	1
The key stage 1 trial	1
1.1 Key stage 1 assessment: background and issues	2
1.1.1 The history of key stage 1 statutory assessment	2
1.1.2 Issues involved in key stage 1 statutory assessment	3
1.1.3 Teacher assessment and key stage 1 assessment	4
1.2 The design of the evaluation	6
1.2.1 Specification and research questions	6
1.2.2 The sample of Local Education Authorities (LEAs) and schools	6
1.2.3 Data collection methodology	9
1.2.4 The representativeness of the sample	14
1.3 Key findings and issues	19
2. The ‘robustness’ of the trial arrangements	21
2.1 A comparative analysis of the data in 2003 and 2004 of Year 2 pupils in the schools in the evaluation sample	23
2.1.1 Comparative analysis of the patterns of performance	23
2.1.2 The effect of LEA on performance	29
2.1.3 The effect of gender on performance	31
2.1.4 The effect of age on performance	32
2.1.5 School type, value-added and PANDA	33
2.1.6 Variation between test / task results and Teacher Assessment	38
2.2 Training and moderation	43
2.2.1 Training	44
2.2.2 Moderation	47
2.3 Other evidence to evaluate the robustness of the data from the trial system	54
2.3.1 Evaluation school data and DfES data - a comparison	54
2.3.1 The confidence of headteachers and LEA assessment coordinators	55
2.4 Key issues and findings	56

3. Workload and management implications of the new arrangements	57
3.1 Workload implications of the new arrangements for teachers, schools and LEAs	57
3.1.1 Pre-trial (2003) experiences and perceptions of workload for teachers and schools in relation to KS1 assessment	58
3.1.2 The impact of the trial (2004) on workload for teachers and schools	59
3.1.3 The impact of the trial on workload in LEAs	66
3.2 Management implications of the new arrangements for schools and LEAs	67
3.2.1 Issues preceding the trial	67
3.2.2 Issues following the trial	69
3.2.3 Overall impact of the new arrangements on school management	75
3.2.4 Impact of the new arrangements on management of assessment within LEAs	76
3.3 Key findings and issues	77
Workload	77
Management	78
4. Views of stakeholders on other issues	80
4.1 The benefits and challenges to schools of the new arrangements	80
4.2 The classroom experience of Year 2 children	83
4.3 Parents' views on reporting of the results	87
4.4 Parents' confidence in teacher assessment	91
4.5 Linking summative and formative purposes for assessment	92
4.6 The use of the assessment outcomes	93
4.7 Comparisons of the new and the old assessment arrangements	95
4.8 Key findings and issues	99

5. Summary of findings and final discussion	100
5.1 Does the new system provide national data for value-added and target setting purposes that are as robust as the current system?	101
Training	103
Moderation	104
5.2 What are the additional workload and management implications of such a system for teachers, schools and LEAs?	104
5.3 The remaining research questions	106
5.3.1 Comparing the new arrangements to the old	106
5.3.2 The emphasis on teacher assessment	107
5.3.3 Parents' views	107
References	110
Appendices	A1
Appendix A	A2
Appendix B	A29
Appendix C	A45
Appendix D	A49
Appendix E	A53

Glossary

AEU	Assessment and Evaluation Unit
AfL	Assessment for Learning
AT	Attainment Target
CPD	Continued Professional Development
CFAS	Centre for Formative Assessment Studies
DfES	Department for Education and Skills
KS1	Key stage 1
KS2	Key stage 2
LEA	Local Education Authority
NAA	National Assessment Agency
NC	National Curriculum
NFER	National Foundation for Educational Research
NLS	National Literacy Strategy
NNS	National Numeracy Strategy
OFSTED	Office for Standards in Education
PANDA	Performance And Assessment
QCA	Qualifications and Curriculum Authority
TES	Times Educational Supplement
TA	Teacher Assessment
W	Working Towards Level 1
Y2	Year 2

1. Introduction: background and design of the evaluation

The key stage 1 trial

A more flexible approach to testing was suggested in the DfES document ‘Excellence and Enjoyment’ (DfES, 2003). The Qualifications and Assessment Authority was therefore asked to trial an approach to testing at key stage 1 which uses national tests to underpin Teacher Assessment, with a Teacher Assessment level being reported nationally. Such a system would offer teachers more flexibility and would place greater emphasis on teacher judgements about children’s progress throughout the year. Overall, it would reduce the amount of testing for 7-year-olds. Part of the aim is to see if the current system can be improved by providing parents with a single set of results, and reducing any possible stress on children.

Local Education Authorities were asked to volunteer for this national trial of the possible new system, and a quarter of all LEAs in England were chosen to take part. These were chosen to be representative of the national profile of LEAs and schools in terms of urban, suburban and rural communities. Introductory information was given to these participating LEAs and a series of national conferences arranged to share experience.

The University of Leeds was awarded the contract to evaluate this trial and report the findings to the QCA by September 2004.

This introduction to the report is divided into 3 sections:

1.1 Key stage 1 Assessment : background and issues

- The history of key stage 1 statutory assessment
- Issues involved in key stage 1 assessment
- Teacher assessment and key stage 1 assessment

1.2 The design of the evaluation

- Specification and research questions
- The sample of LEAs and schools
- Data collection methodology
- The representativeness of the sample

1.3 Key findings and issues

Note

Throughout the report it should be noted that there is an occasional ambiguity in the use of the phrase “teacher assessment”, which needs to be considered when interpreting the accounts of teachers and authors.

In the new arrangements the phrase “Teacher Assessment” is used for the final reported judgement that takes task and test results into account. However, in the existing arrangements for KS1 assessment, Teacher Assessment is reported separately from test and task results. For some teachers and authors, “teacher assessment” means what the teacher judges the child’s achievement to be *independently of* test or task outcomes, as a result of monitoring classwork. The process of arriving at a judgement to be reported in the new arrangements may be thought of as firstly evaluating the children’s achievements on the basis of work in class (i.e. doing some teacher assessment), and then considering that teacher assessment alongside the test and task result, to arrive at a final judgement – which is then called the Teacher Assessment (note the capital letters).

Evaluation of the trial assessment arrangements for key stage 1

As a result of this, when a teacher or author talks of “teacher assessment” it is sometimes difficult to tell if they mean the final judgement that takes task and test into account, or the judgement of achievement based only on work in class. When it is clear that what is meant is the reported level, then capital letters will be used (Teacher Assessment, which may be abbreviated to TA in tables), but the ambiguity in other cases should be kept in mind.

1.1 Key stage 1 assessment: background and issues

- 1.1.1 The history of key stage 1 statutory assessment
- 1.1.2 Issues involved in key stage 1 assessment
- 1.1.3 Teacher assessment and key stage 1 assessment

1.1.1 The history of key stage 1 statutory assessment

The first KS1 tests were administered in **1991**, to assess children in Mathematics, English and Science. They were mostly task based and for English included reading text aloud and an optional comprehension test for levels 2-3. Maths and science were task based with some worksheets. In addition, teachers also had to complete their own Teacher Assessments (TA). There was controversy over the amount of time involved in these assessments for teachers and how much teaching time these 7 year olds were missing (Shorrocks-Taylor, 1999). There was also great variability in training for the Teacher Assessments. Another concern about the 1991 tests was the low level of agreement between the TA and the test scores, even ones measuring basic skills (Shorrocks et al, 1991, ENCA).

As a direct result of many of these expressed concerns, the 1992 tests were streamlined, and for the next few years little significant change occurred year-on-year, but there was a gradual move towards more pencil and paper assessments and fewer tasks. The most significant changes were:

- **1993** the Dearing review recommended that the TA should be reported to parents separately;
- **1994** the science tests were abandoned and, in this subject, TA was used for all levels including level 4;
- **1995** 7-year-olds who were thought to be working at level 4 were tested using the KS2 materials in mathematics and English, although some thought these materials were inappropriate for KS1 children;
- **1996** the tasks could be administered from January to provide even longer for teachers to complete them and the tests were given in May; again level 4 was tested using KS2 materials;
- supply cover funding became available for all subjects to cover costs of staff time
- NFER’s evaluation of KS1 testing said overall it was functioning ‘reasonably satisfactorily’ (Sizmur and Burley, 1997)

By **1998**, the KS1 assessment regime had become more test based to make it more manageable and the workload more balanced, but the introduction of a comprehension test in 1997 again meant there were workload issues. The move towards paper tests “*undoubtedly improves manageability for the teachers but may have less validity in educational terms for the pupils of this age*” (Shorrocks-Taylor, 1999, p.116).

In **2003** the KS1 tests were changed again. The main changes were the use of a spelling test with *all* children, handwriting being part of the writing test, and a significantly changed mark scheme for a longer writing task. The other main change was the separate level 3 test in mathematics. The longer writing task and the fact that at least part of it had to be a story was controversial as it gave less freedom for teachers to choose the type of writing a child might do which might help emphasise that

Evaluation of the trial assessment arrangements for key stage 1

child's strengths. There were also workload issues again due to the complexity of the marking schemes, (News & Opinion, TES, 16/05/03).

The national KS1 assessments for **2004** (excluding those LEAs in the trial of the new system) include the following:

English

Reading task (level 1-2)

Reading test (level 2)

Reading test (level 3)

Writing tasks (levels 1-3):

long task

short task

spelling test

handwriting judgement

Mathematics

Maths task (level 1)

Maths test (level 2)

Maths test (level 3).

(www.qca.org.uk/ages3-14/)

1.1.2 Issues involved in key stage 1 statutory assessment

It is important to note that tests and tasks are not without their problems. There are potential problems with the tasks, in part due to teachers trying to glean information about individual children within a small group setting. Children may work at different rates, and trying to get the cognitive demand the same for children while using different questions in a small group can be difficult and also very time consuming. Teachers usually score on a mark sheet during the session. All this can lead to unreliability in task assessment (Shorrocks-Taylor, 1999) and this may be why so much of the KS1 assessment has become test-based rather than tasks. These issues still exist, especially for children working at around level 1, who still complete the tasks, and for whom tasks may well be a more reliable form of assessment at KS1.

However, formal tests for seven-year-olds are also not without their problems. The formality and individualised working of the test situation may prove daunting and stressful for young children, unless careful preparations has been made. As in any test situation, performance on the day may be affected by many influences, and will represent skills and knowledge assessed at that one point in time. This may or may not be a true reflection of the child's skills and understandings. A comparatively short test may also not be able fully to address the breadth of the curriculum that has been delivered.

However, a recent CFAS (Centre for Formative Assessment Studies, 2002) report found that many teachers did think the overall tests and tasks were suitable for KS1 and that they did use the results to help inform their Teacher Assessment. However, teachers thought that some issues could be addressed in order to improve the relationship between Teacher Assessment, tests and tasks at KS1. These included:

- greater parental understanding that Teacher Assessment and national test results will not always match;
- more understanding that Teacher Assessment is built upon wider knowledge of the child;
- a consideration of the appropriateness of testing for 7-year-olds;
- the need for more guidelines for Teacher Assessment, especially for science.

Evaluation of the trial assessment arrangements for key stage 1

The findings of CFAS seemed to reveal a general trend of teachers and schools, on the whole, coping well with statutory assessments. The Annual Survey of Trends in Education (Spring 2003, digest No 14) carried out by the NFER found that headteachers, at least, were much more concerned about budgets, staffing and special educational needs and OFSTED inspections rather than national assessment, although the concern about these assessments has risen from the previous year and 32% did mention it. Another relevant point from this report suggests that 98% of headteachers felt that staff workload and the administrative burden on teachers was an issue for major concern. Assessment itself was not mentioned by any of the heads as of being of particular concern to their governing bodies.

1.1.3 Teacher assessment and key stage 1 assessment

Teachers are expected to assess their pupils regularly and keep records of their progress in relation to the National Curriculum. However, assessment does not just occur for one purpose within a classroom and much of it is informal: it is used to see if children are ready to move onto another activity or to observe whether and how they are learning. Rea-Dickins (2001) suggests that classroom assessment can have different identities. These are:

- Bureaucratic – this reflects an external identity where there is an obligation on the part of schools and teachers to provide information, such as NC levels.
- Pedagogic – this has an internal identity, it is required by the teacher to support and gather information about individual learners.
- Learning – this focuses upon learning through assessment and the role of the learner in this, it develops learner awareness and motivates them.

In order to make the teacher assessment judgements within statutory assessments, formative assessments are supposed to be carried out and records kept, but this demands careful planning and moderation. It has also been the case that, historically, Teacher Assessment has often been seen to have secondary importance to tests. It is used, in part, to assess some of the curriculum which are not covered in the tests.

At key stage 1, Teacher Assessment started off as a central issue but never had the same profile or recognition as the tests (Gipps, 1995), yet some have suggested it might indeed be the best way to assess young children as an integral part of formative assessment (Harlen 1992).

When the assessment arrangements for the National Curriculum were being designed this combination of formative and summative purpose was built into them. The report of the Task Group on Assessment and Testing (TGAT, 1988), on which the arrangements were to be based, stressed that assessment should provide information on all pupils' strengths and weaknesses in relation to their progression through the National Curriculum as a whole, which teachers could then use in planning further learning. It argued that the basis of the national assessment system should be essentially "formative" (TGAT, 1988, para. 27) and this would be best accomplished through Teacher Assessment with externally devised Standard Assessment Tasks (SATs) to moderate Teacher Assessment and ensure some degree of comparability in the system. At the end of the key stages (years 2, 6 and 9) grades were to be produced by combining Teacher Assessment grades with SAT results. It could be argued that "teacher assessment might have produced better and faster gains in performance than the present emphasis on using testing to raise standards." (Shorrocks-Taylor, 1999, p.172). In 2004, England is the only one of the four home nations to have formal tests for 7 year olds, the Welsh Assembly having recently dropped KS1 testing in favour of teacher assessment.

Teacher assessment though is not without its problems. A recent study by Hall & Harding (2002) looked at the assessment practice at KS1 in 6 schools over a period of 2 years. They were studying the implementation of level descriptors in teacher assessment; how their interpretation and application

Evaluation of the trial assessment arrangements for key stage 1

might not be straightforward. The Level descriptors (LD) were then used to provide a Teacher Assessment level. There are problems with these level descriptors, since they assume that all teachers interpret them in the same way, yet they are very complex (William, 1998, Sainsbury and Sizmur, 1998).

Hall and Harding's study looked at whether schools engaged in what they call a "community of assessment practice", and found two distinct approaches to this assessment. One was *collaborative* and the other *individualistic*. The collaborative school did have assessment discussions and tried to get a shared meaning for assessment. The head organised time out of lessons for discussions and shared results and portfolios with parents and sometimes children. The Year 3 teachers took more notice of results they were given from Year 2 because they hadn't occurred in isolation. On the other hand, in the individualistic school teachers worked alone to provide a level, and the results were taken less seriously by the Year 3 teachers and were sometimes not shown to parents. Overall they suggest that "*there is evidence of an assessment community of practice within some schools, but such communities are confined mainly to the teachers within those particular schools*" p.11. There was a lack of networking with nearby schools or involvement of parents and children.

Some very recent evidence (Chisholm, 2004) has also highlighted the fact that the development, implementation and permeation of whole-school assessment policy has declined in importance in schools since the mid-1990s. If this is the case, then one of the key means of achieving the 'community of assessment practice' in schools has been undermined.

Hall and Harding also found that there was less discussion about teacher assessment in the second year of their study, due to the introduction of the National Literacy Strategy, and that funding was not available for teachers to moderate their teacher assessment results. "*The real danger now in relation to TA is that not only are such cultures or communities of assessment practice not being supported and developed, but the ground that was gained in the early and mid-1990's could be receding.*" (Hall and Harding, 2002, p.13)

However, even though this lack of assessment community may be an issue there may be few problems in terms of the comparisons between tests scores and Teacher Assessment. A recent study suggests that there is little or no gap between Teacher Assessment levels and test levels. Durant (2003) looked at KS1, 2 and 3 results over 5 years from 1998-2002. This included 32,000 children for KS1 from Worcestershire. Durant found a high level of agreement between test levels and Teacher Assessments, for Worcestershire schools.

He found that for reading, writing, and maths 50% of test and Teacher Assessment were identical and most were within 1 test level. This was a fairly consistent finding for reading and writing from 1998-2002. For maths however, even though the proportion of cases was consistent at 85%, the proportion showing a difference rose from 14% in 1997/98 to 22% in 2001/02 and a corresponding difference in those scored higher by the Teacher Assessment from 21% down to 12%.

The issue of the stress caused to young pupils has also been highlighted over the years, although published and verifiable evidence is hard to find. Connor (2003) followed up a study done in 2001 which highlighted risk of stress among children taking the national tests at ages 7 and 11. It was possible to identify a small group of children with clear symptoms of stress during the round of tests. This may therefore need to be considered as part of the wider decision-making.

The issue of value-added is also an important one, given that the key stage 1 results play a significant role in evaluating the national picture of progress in schools. This is not the place for a technical discussion, except to mention the fact that a recent report from the NAHT (National Association of Head Teachers) prepared by Curriculum, Evaluation and Management Centre at the University of Durham states value-added scores in Primary School League Tables may be misleading in their current form. They argue that a number of issues need urgent attention at key stages 1 and 2, among the most significant being:

size of school – small schools results fluctuate markedly, for obvious reasons;

quality of KSI data – internal marking makes validity of comparisons between schools difficult;

turnover – present measures do not account for pupils entering and leaving schools, and in some, as many as half the pupils in Year 6 might not have been at school 4 years before.

1.2 The design of the evaluation

1.2.1 Specification and research questions

1.2.2 The sample of LEAs and schools

1.2.3 Data collection methodology

1.2.4 The representativeness of the sample.

1.2.1 Specification and research questions

It is clear from the above analysis that assessing 7-year-olds has been a controversial issue here since the earliest days of the implementation of the national testing programme (Shorrocks *et al*, 1991; Shorrocks-Taylor 1999). Leaving aside the matter of the very vulnerable age of these pupils, the educational and technical debate has been about balancing the requirements of **reliability**, **validity** and **manageability** in the summative key stage assessments. The trial arrangements in 2004 are further addressing these problems by the use of testing to underpin Teacher Assessment, and the evaluation of the trial is focused on these three requirements.

This translates into the following research questions, which were contained in the specification for the evaluation:

- **Does the new system provide national data for value-added and target setting purposes that are as robust as the current system?**
- **What are the additional management and workload implications of such a system for teachers, schools and LEAs?**
- **What is the impact on, and views of parents, children and governors, as well as those of the school/LEA stakeholders of the new system compared to the existing one?**
- **How valuable do schools and LEAs find the emphasis on Teacher Assessment in comparison to the previous system?**
- **Do parents find the new system easier to understand and does it give them enough information?**

Findings on each of these questions are reported in the subsequent chapters.

1.2.2 The sample of Local Education Authorities (LEAs) and schools

At the start of the project information was provided by the QCA about how LEAs currently audit and subsequently plan to moderate key stage 1 assessments, their guidance to schools about assessment and reporting, and their intentions for implementing the trial.

This information was used to decide the sample of LEAs for the evaluation. In addition, information about the 2003 Teacher Assessment and test and task data for the children in the LEAs in the trial was obtained from the DfES, which holds the relevant data centrally. These data were used initially for two purposes, firstly to choose the LEAs and the schools within those LEAs for the evaluation, and

Evaluation of the trial assessment arrangements for key stage 1

secondly to ascertain whether these LEAs and schools are representative of the rest of the LEAs in the trial. It will also be used in Chapter 2 for comparison purposes with the data that has been produced by the trial.

The trial of the new arrangements involved 34 LEAs across England, as shown in Table 1.1.

Table 1.1 The 34 LEAs in the national trial, and number of pupils per LEA for whom there was some data for KS1 assessment in 2003

LEA	No of pupils
Birmingham	13720
Bromley	3464
Cornwall	5438
Coventry	3706
Derbyshire	8627
Dorset	4084
Durham	5640
East Riding of Yorkshire	3405
Essex	15576
Halton	1380
Haringey	2837
Harrow	2375
Hertfordshire	12744
Kent	16178
Kingston upon Thames	1586
Kirklees	5010
Leicestershire	7147
Lewisham	2537
Liverpool	5309
North Lincolnshire	1861
North East Lincolnshire	1980
North Somerset	2107
Oxfordshire	6401
Portsmouth	2006
Sheffield	5699
South Tyneside	1787
Southend-on-Sea	2104
Staffordshire	9537
Tower Hamlets	2795
Wakefield	3848
Waltham Forest	2772
Wigan	3616
Wiltshire	5187
Wolverhampton	3013
Total	175476

The intention was to work with 10 different trial LEAs for the evaluation. Initially 12 LEAs were approached. Of these, 8 agreed to participate in the evaluation. Since a larger sample was required, other LEAs were approached which were as similar as possible to those who had declined involvement. As a result a sample of 10 trial LEAs were brought together for the focused information gathering for the evaluation.

Table 1.2 The ten LEAs in the evaluation, and number of pupils per LEA for whom there was some data for KS1 assessment in 2003

LEA	No of pupils
Dorset	4084
East Riding of Yorkshire	3405
Essex	15576
Haringey	2837
Harrow	2375
Hertfordshire	12744
Leicestershire	7147
South Tyneside	1787
Wigan	3616
Wolverhampton	3013
Total	56584

The focused sample targeted 150 schools in the 10 evaluation LEAs. These 150 schools were selected by the LEA coordinators. A list of 15 schools in each LEA was asked for, to include a representative sample in terms of school type and size, 10 of which were to be schools that would undergo moderation. These schools were invited to participate in the evaluation of the trial.

Responses from schools varied. In one LEA, all 15 schools agreed to be involved, whilst in others the 'yes' response was only 5 of the 15. Where the agreement to be involved numbered less than 10 schools in any LEA, the LEA coordinator was approached for more schools, who were written to ask for their involvement. The sample was finalised at the end of April 2004 and consisted of 107 schools in 10 LEAs. Table 1.3 shows the number of schools in each LEA involved in the evaluation, and the number of Year 2 pupils for whom assessment data were collected in these schools last year.

Table 1.3 Number of evaluation schools and pupils in the 10 evaluation LEAs

LEA	No. of schools	No of pupils
Dorset	14	620
East Riding of Yorkshire	12	378
Essex	13	439
Haringey	9	513
Harrow	8	525
Hertfordshire	8	257
Leicestershire	10	400
South Tyneside	10	324
Wigan	10	321
Wolverhampton	13	513
Total	107	4290

The sample size of 4290 children gives a minimum level of accuracy for point estimation of $\pm 1\%$ at the 5% confidence interval.

The size of the evaluation LEAs in comparison to all the LEAs in the trial is shown in Figure 1.1.

Figure 1.1 Trial and evaluation LEAs in order of size (number of pupils)

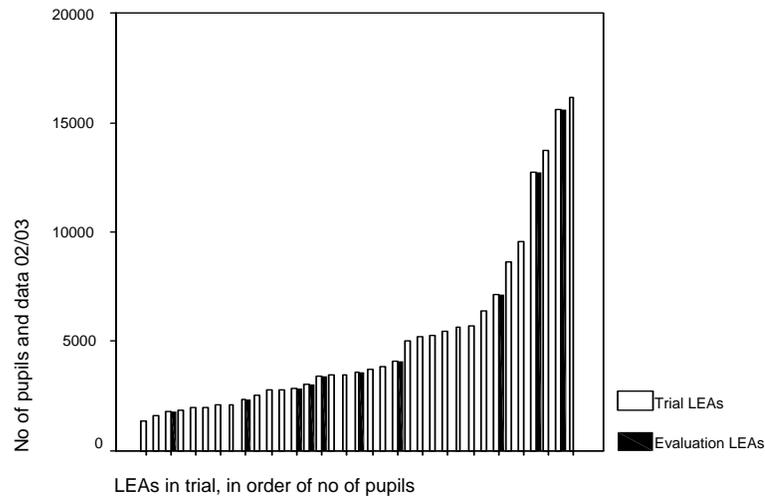


Figure 1.1 shows that the ten LEAs in the evaluation range in size in a similar way to the 34 in the trial. There are small, medium-sized and large LEAs represented in the evaluation, giving it good size representativeness. The detail of representativeness in other dimensions is given in a later section of this introduction.

1.2.3 Data collection methodology

Questionnaires

The early phase of the evaluation included the development of LEA, headteacher, teacher, junior school, parent and governor questionnaires, interview schedules and focus group prompts, and their trialling. As part of the quality control of the questionnaires the advice of an expert in research methods with many years of experience in compiling and administering questionnaires was sought. After modification, draft questionnaires were sent to QCA for comment, and further modifications implemented. At this point, the questionnaires were sent to a number of heads, teachers and LEA personnel in local LEAs that are in the trial but not in the evaluation, to seek further feedback. Final modifications were made in the light of these comments. The final questionnaires are reproduced in Appendix A. The questionnaires have both closed questions, with options to be identified, the responses to which were analysed quantitatively, and open-ended questions asking for comments and explanations, which were analysed qualitatively. The results for responses to the closed questions are reported in Appendix B.

Interviews and Focus Groups

In addition interview and focus group schedules were developed in a similar way, and trialled appropriately. These are shown in Appendices C and D respectively.

The timing and content of data collection

The questionnaires were distributed in 2 stages: in April 2004, the first set of questionnaires were sent to headteachers, teachers, junior schools, governors and LEA personnel. The sample for the questionnaires was all the relevant personnel in the sample of 10 LEAs, and all the coordinators in the 34 LEAs in the trial, the teachers of all 107 schools in the sample (185 teachers and 107 head teachers), and the governors of the schools in the evaluation.

Evaluation of the trial assessment arrangements for key stage 1

The second stage of questionnaire data collection took place in June 2004, when questionnaires were sent to headteachers, teachers, LEA coordinators and parents. The sample for the questionnaires was all the LEA coordinators in the 34 trial LEAs, the teachers of all 107 schools in the sample (185 teachers and 107 head teachers), and all the parents of the Year 2 children in the sample schools.

1) Headteacher questionnaires 1 and 2

The first headteacher questionnaire, issued in April, was to survey views about the key stage 1 assessment arrangements for 2002/3, including matters of workload, support, and value, and also expectations about the forthcoming trial assessment arrangements, focusing on benefits, challenges and workload.

The second headteacher questionnaire, issued in late June, collected views about the experienced assessment arrangements, especially the aspects of workload, support, value and levels of confidence.

2) Teacher questionnaires 1 and 2

The first teacher questionnaire, issued in April, reviewed teachers' views about the assessment arrangements for 2002/3, including the workload, support and value, and the extent to which test and task results had been used to inform Teacher Assessments. It also explored the assessment arrangements of the trial, including expectations of workload and benefits, and intentions for keeping examples of children's work and attending LEA training meetings.

The second teacher questionnaire, issued in late June, obtained views about the workload in the new arrangements, the support given, the perceived value of the assessment system, and the levels of confidence in the judgements made, and also elicited information about the children's work that was kept, the actual tests and tasks that were used, and the degree of variation between test and task result and Teacher Assessments.

3) LEA questionnaires 1 and 2

The first questionnaire to LEA assessment coordinators, issued in April, investigated expectations about the system being trialled, asking for views about workload, benefits and challenges, and LEA systems of moderation and value-added analysis.

The second LEA coordinator questionnaire, issued in late June, looked at the issues of workload, challenges, benefits and confidence in the light of the experience of implementing the new arrangements.

4) LEA - other staff

Questionnaires were also issued to other LEA personnel to ascertain the involvement of different interest groups in the implementation of the key stage 1 assessment arrangements, and to obtain views about the relationship between the new system and other initiatives being implemented in schools. The targeted personnel were the person or persons responsible for implementing the Numeracy and Literacy aspects of the Primary National Strategy, the English and Mathematics Advisors (if different) and the person with a responsibility for the Early Years.

5) Governor questionnaires

The questionnaire to school governors, issued via the schools in April, was administered to find out the extent that governors had been made aware of the new assessment arrangements, and their views about possible benefits, challenges and the impact on workloads.

6) Junior school questionnaires

The questionnaire to Junior Schools, distributed in April, attempted to discern possible differences between the use of assessment results from feeder infant schools under the old and new arrangements, by asking about what the schools had done last year and about what they intended to do this year.

7) Parents questionnaires

The questionnaire to parents, sent out through the schools in late June / early July, was used to determine the extent parents were aware of the new arrangements for assessment in Year 2, what their views were about the accuracy of the results, whether they asked for test and task results in addition to the results they received, and what the children's attitudes were to the process and the outcomes.

A package, comprising head teacher questionnaires, teacher questionnaires and governor questionnaires was sent to all participating schools in April 2004. All questionnaires were sent with a separate FREEPOST return envelope. The governor questionnaires were also accompanied by an information sheet explaining the trial (produced by QCA and available on their website). Schools were asked to return the head/teacher questionnaires by the end of the first week of May and for the governor's questionnaires to be returned within two weeks of them having been distributed to appropriate colleagues. Junior schools were sent their questionnaires at this time also, again with FREEPOST return envelopes.

All trial LEAs were sent the LEA coordinator questionnaires in April 2004, again with return FREEPOST envelopes. The evaluation LEAs were also sent questionnaires for personnel involved in Early Years, Literacy, Numeracy and those involved in advisory positions in Mathematics and English. Again FREEPOST return envelopes were provided.

All the questionnaires to heads/teachers, governors and junior schools had the name of the LEA on, but not the name of the school. All the questionnaires to the LEAs had no identifying names. All questionnaires to parents had the name of the LEA on and were colour-coded by school. This latter identifier was included so that it would be possible to ascertain how many schools did not return any parent questionnaires. It was clear from this that for a number of schools they were not administered.

The second tranche of questionnaires was sent out in June, in the same way with FREEPOST envelopes. Parent questionnaires were sent to the school for them to distribute via their Year 2 children. All parent questionnaires were in an envelope with a FREEPOST envelope for return.

Table 1.4 summarises the numbers of questionnaires sent out and received back completed.

Table 1.4 Number of each type of questionnaire sent out and number received back completed

Questionnaire	Number of questionnaires administered	Number of completed questionnaires received ¹
LEA co-ordinator Q1	34	26
LEA co-ordinator Q2	34	22
LEA other	30	22
Headteacher Q1	107	83
Headteacher Q2	107	63
Teacher Q1	185	115
Teacher Q2	185	88
Junior schools	41	27
Governors	1620	162
Parents	One to every parent of a Year 2 child in an evaluation schools, approximately 5000 ²	570

Interviews and Focus groups

A combination of interviews and focus groups was also used to access additional detail, and to reinforce the validity of the questionnaire data. The interviews were utilised mainly to access processes and practices, whereas focus groups were utilised mainly to access views and beliefs.

An embedded design (Yin, 1994) has been used for the interviews and focus groups in order to investigate the overlapping perspectives of the implementation of the trial among teachers, governors, parents and LEA personnel. Such a design assists in the processes of intercollation in data analysis and has helped to provide broader perspectives.

Focus groups were set up, involving parents in selected schools from the sample LEAs. There were also focus groups of teachers from evaluation schools and some with teachers from non-evaluation schools. Interviews were conducted with key LEA personnel in all the 10 LEAs of the sample, and with teachers from selected schools in those LEAs, with some schools being those in the evaluation and others being trial but not evaluation schools.

Focus Groups

1) Teachers

The focus groups for teachers, held in late June and early July, followed up in greater depth the issues explored in the questionnaires, both matters which the first questionnaires had raised, and also elements that could not be pursued in detail in the questionnaire context.

2) Parents

The focus groups for parents, also held in late June and early July, were also to explore in greater depth the issues covered by the parent questionnaires, attempting both to validate the opinions found in the questionnaires and to add detail and meaning to the views expressed.

¹The numbers given here are the numbers returned in time for analysis. For all types of questionnaire a few additional returns were received after the analysis had taken place.

²However, the school distributed the questionnaires and it seems that some schools did not want parents to have these questionnaires and did not distribute them.

Interviews

1) Teachers

The interviews with teachers, held in late June and early July, followed up in greater depth the issues explored in the questionnaires. Those with teachers in evaluation schools focused on elements that could not be pursued in detail in the questionnaire context. The interviews with teachers in non-evaluation schools also asked questions about workload, in order to ascertain the effect of the evaluation itself on teachers' perceptions of the workload demands of the trial.

2) LEA personnel

The interviews with LEA coordinators, held in late June and early July, also explored in greater depth the important issues of concern to the evaluation than that possible in questionnaires.

All interviews and focus groups were minidisc-recorded, summarised and transcribed.

Table 1.5 Summary of focus groups and interviews

Events	Numbers involved
Teacher interviews	22
LEA interviews	9
Teacher focus groups	10
Parent focus groups	6

In those LEAS where so few parents volunteered (in one case none volunteered) those who did were given the AEU phone number to give any comments. One parent did this and their comments were incorporated into the analysis.

Additional data collection activities

1) Training session

Training sessions for Year 2 teachers were observed in January/February, and moderation sessions were observed in June.

2) Data collection and the pupil information form (PIF)

Detailed information about Teacher Assessment, test and task results for the evaluation schools was requested from the 107 schools involved in the evaluation. Participating teachers were asked to provide performance data and pupil information for all their Year 2 children. The performance data comprised Teacher Assessment and task/test level in English and Mathematics. The pupil information comprised gender and date of birth. Teachers were given the choice of how this information would be submitted: hard copy or electronically. If electronically a number of options were suggested, including Word, Excel or ascii data file. Where schools did not return their data, the data set was supplemented with DfES data that had information about Teacher Assessment but not test/task. Using both sources data on 101 of the 107 evaluation schools were obtained.

This data set, the 2003 DfES data set (of Teacher Assessment, task and test results) and a 2004 data set provided by DfES (with Teacher Assessment only) have been used to carry out the statistical analysis reported later in this chapter and in Chapter 2.

Evaluation of the trial assessment arrangements for key stage 1

The timeline for the work of the evaluation is given in Figure 1.2 below.

Fig 1.2 Timeline of the research

Activities	Dates	Phase 1		Phase 2		Phase 3	
		October-December 2003	January - March 2004	April 2004	May-June 2004	June-July 2004	August 2004
Information gathering from QCA, LEAs and DfES							
Selection of sample							
Questionnaires, interview schedules and focus group							
Administration of first stage of questionnaires							
Administration of second stage of questionnaires							
Interviews and focus groups							
Data analysis (quantitative and qualitative)			Phase 1 data			Phase 2 data	
Report writing							

This research design aimed to produce a rigorous analysis of the pilot system under evaluation;

- an evaluation of whether the new system is fit for purpose (reported in Chapter 2);
- an evaluation of workload and manageability implications of the new system (reported in Chapter 3);
- an evaluation of the views of LEA personnel, headteachers, teachers, governors and parents about the trial arrangements (reported in Chapter 4);

1.2.4 The representativeness of the sample

It is important that the LEAs in the evaluation and, in turn, the evaluation schools are representative of the sample of 34 LEAs in the trial, so that the findings and conclusions from the 10 LEAs and 107 schools within them can be used to make inferences about the trial overall. Some information has already been provided earlier in this introduction about the range, number and sizes of the LEAs and schools selected for this evaluation. This sub-section focuses on additional measures of representativeness: firstly a comparison of proportions of children attaining different levels, and secondly a comparison of mean level score.

Proportions of children attaining different levels

Tables 1.6-1.8 below show, for 2003 data, the number and percentage of each sample (all trial schools, all evaluation LEA schools, and evaluation schools) attaining each level in the reading task/test, the writing task, and the mathematics tests/task. The accompanying Figures (1.3-1.5) give the percentage information visually.

Evaluation of the trial assessment arrangements for key stage 1

Table 1.6 Levels achieved in the Reading task/test

	Trial LEAs		Evaluation LEAs		Evaluation schools	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
W	5217	3.0	146	2.6	103	2.4
1	21358	12.2	6267	11.1	493	11.6
2C	17051	9.8	5248	9.3	351	8.2
2B	31725	8.2	9908	17.6	764	17.9
2A	50295	28.8	16390	29.1	1248	29.2
3	48878	28.0	17036	30.3	1309	30.7
Total	174524	100.0	56295	100.0	4268	100.0

Figure 1.3 Percentage of each level achieved in Reading task/test

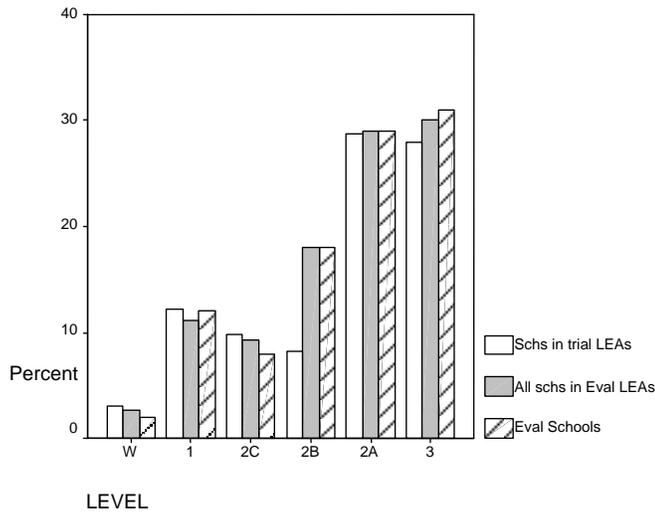


Table 1.7 Levels achieved in the Writing task

	Trial LEAs		Evaluation LEAs		Evaluation schools	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
W	8956	5.1	2342	4.5	188	4.4
1	22847	13.1	6949	12.3	520	12.2
2C	33708	19.3	10228	18.1	770	18.0
2B	42757	24.5	13724	24.4	1026	24.0
2A	38111	21.8	12288	21.8	926	21.7
3	28111	16.1	9759	18.4	838	19.5
Total	174490	100.0	55290	100.0	4268	100.0

Figure 1.4 Percentage of each level achieved in Writing task

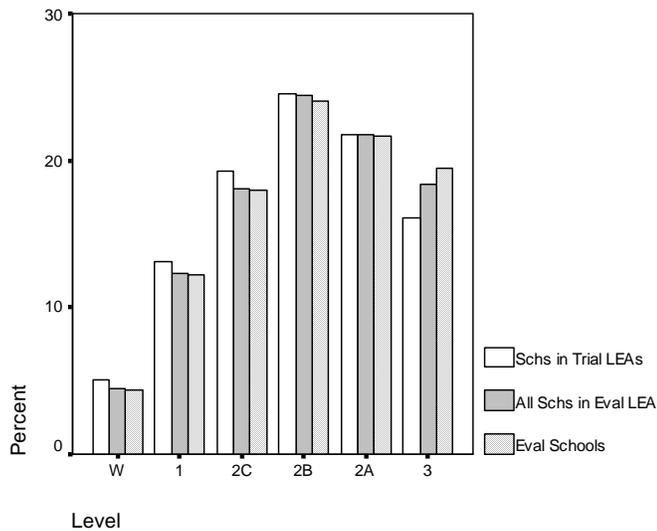
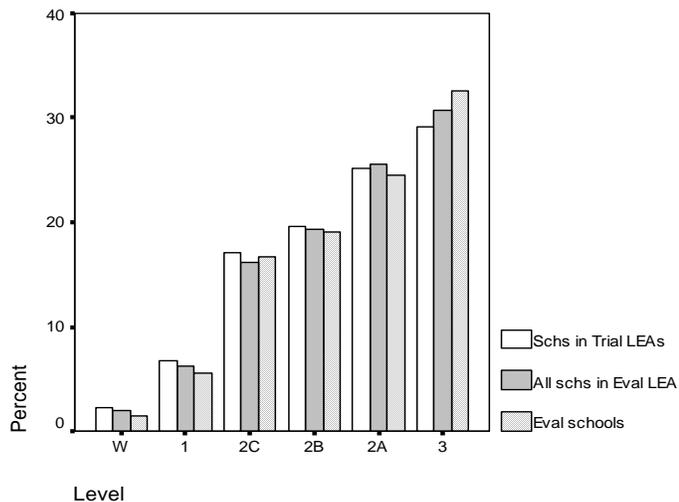


Table 1.8 Levels achieved in the Mathematics test or task

	Trial LEAs		Evaluation LEAs		Evaluation schools	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
W	3882	2.2	1119	2.0	63	1.5
1	11837	6.8	3473	6.2	238	5.6
2C	29776	17.1	9100	16.2	712	16.7
2B	34244	19.6	10880	19.3	817	19.1
2A	43812	25.1	14380	25.5	1046	24.5
3	50987	29.2	17352	30.7	1392	32.6
Total	174538	100.0	56304	100.0	4268	100.0

Figure 1.5 Percentage of each level in the Mathematics test or task



Evaluation of the trial assessment arrangements for key stage 1

In each case shown in Tables 1.6-1.8 and Figures 1.3-1.5, the proportions of children in the evaluation sample and in the evaluation LEAs that attained different levels in 2003 are very similar in all subjects to the children in all the schools of the trial. This indicates the evaluation sample is representative.

Mean level score

The representativeness of the evaluation LEAs and schools can also be judged by comparing performance statistically, comparing LEAs in the evaluation with those in the trial but not in the evaluation (Table 1.12) and then comparing schools in the evaluation with those in the same LEA that are not in the evaluation (Table 1.13). These comparisons are shown schematically below.

Fig 1.6 Comparative data in Table 1.10

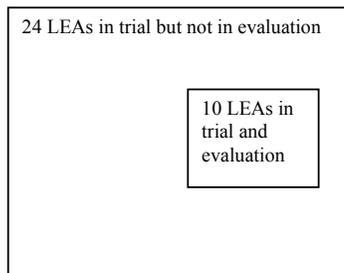
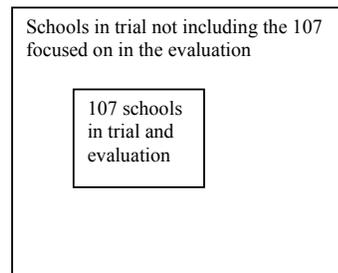


Fig 1.7 Comparative data in Table 1.11



The levels were scaled in the following way to be able to make a comparison of means.

Table 1.9 Scale conversion used to compare means

Level	Scale
W	1
1	3
2C	5
2/2B	6
2A	7
3	9
4	12

Table 1.10 compares the mean 2003 KS1 results of the Year 2 children in the 10 evaluation LEAs to the Year 2 children in the 24 LEAs that are in the trial but not in the evaluation, across all relevant subjects, using the conversion scale shown in Table 1.9.

Table 1.10 Mean performance at LEA level of 2 groups, LEAs in the evaluation and LEAs in the trial but not in the evaluation, 2003 data (scaled)

Subject		N	Mean	Std. Dev
Speaking & Listening TA	Not in evaluation	118340	6.2477	1.898
	In evaluation	56367	6.3453	1.844
	Total	174707	6.2792	1.881
Reading TA	Not in evaluation	118345	6.2793	2.086
	In evaluation	56368	6.4306	2.041
	Total	174713	6.3281	2.072
Writing TA	Not in evaluation	118344	5.8243	1.931
	In evaluation	56366	5.9600	1.898
	Total	174710	5.8681	1.922
Reading Task	Not in evaluation	86378	5.3332	1.648
	In evaluation	39255	5.4266	1.596
	Total	125633	5.3623	1.633
Reading Test	Not in evaluation	97807	7.1206	1.449
	In evaluation	47936	7.2252	1.452
	Total	145743	7.1550	1.451
Writing Task	Not in evaluation	118202	5.8019	2.074
	In evaluation	56288	5.9792	2.070
	Total	174490	5.8591	2.074
Maths TA	Not in evaluation	118345	6.3739	1.874
	In evaluation	56369	6.4896	1.840
	Total	174714	6.4113	1.864
Maths Task & Test	Not in evaluation	118234	6.5984	1.961
	In evaluation	56304	6.7339	1.925
	Total	174538	6.6421	1.951

There were differences between LEAs in the evaluation and those not in the evaluation, in every subject and for both Teacher Assessment and test/task. In each case the mean of the scaled scores was higher in the evaluation LEAs. However, these differences are small, ranging between 0.08 and 0.18 on the scale (less than one fifteenth of a level).

Therefore the sample of ten LEAs may be considered reasonably representative of the 34 in the trial.

Table 1.11 compares the mean 2003 KS1 results of the Year 2 children in the 107 evaluation schools within the 10 evaluation LEAs to the Year 2 children in all the other schools in the trial, across all relevant subjects, using the conversion scale shown in Table 1.9.

Table 1.11 Mean performance at school level of 2 groups, schools in evaluation compared to those in the trial but not in the evaluation, 2003 data (scaled)

		N	Mean	Std. Dev
Speaking & Listening TA	Not in evaluation	170393	6.2747	1.881
	In evaluation	4314	6.4536	1.872
	Total	174707	6.2792	1.881
Reading TA	Not in evaluation	170399	6.3235	2.071
	In evaluation	4314	6.5118	2.084
	Total	174713	6.3281	2.072
Writing TA	Not in evaluation	170396	5.8647	1.921
	In evaluation	4314	6.0009	1.928
	Total	174710	5.8681	1.922
Reading Task	Not in evaluation	122645	5.3599	1.633
	In evaluation	2988	5.4625	1.612
	Total	125633	5.3623	1.633
Reading Test	Not in evaluation	142062	7.1530	1.451
	In evaluation	3681	7.2315	1.461
	Total	145743	7.1550	1.451
Writing Task	Not in evaluation	170173	5.8543	2.071
	In evaluation	4317	6.0479	2.090
	Total	174490	5.8591	2.074
Maths TA	Not in evaluation	170400	6.4070	1.864
	In evaluation	4314	6.5786	1.831
	Total	174714	6.4113	1.864
Maths Task & Test	Not in evaluation	170221	6.6377	1.952
	In evaluation	4317	6.8175	1.885
	Total	174538	6.6421	1.951

There were differences between the performance of the children in the schools in the evaluation and those in other schools in every subject, and for both Teacher Assessment and test / task. Again the scores of the children in the evaluation sample were consistently higher, but again only by a small amount, ranging from 0.08 to just under 0.2 on the scale (i.e. no more than one fifteenth of a level).

The evaluation sample of 107 schools may be considered reasonably representative of the trial sample.

1.3 Key findings and issues

In relation to the background to the key stage 1 statutory assessments:

- KS1 assessments have varied historically, with some aspect changing almost every year. They are now much more pencil and paper orientated than when they were first introduced, and a different balance has been struck between reliability, validity and manageability in the assessments.
- On the whole teachers are happy with the quality and style of the tests/tasks, although the published research shows that some still think it is inappropriate to formally test 7 year olds, and there may be some issues concerning the stress caused to young pupils.
- Over the years, many points have been raised about Teacher Assessment, and these are undoubtedly complex, especially in terms of the relationship, theoretical and practical, between this form of assessment and the national test results.

Evaluation of the trial assessment arrangements for key stage 1

- Some schools do have a successful and supporting assessment community (but not all), although this does not usually extend across schools (and probably not across LEAs), which would need to be the case for Teacher Assessment to be reliable and robust.
- Evidence from Worcestershire suggests that, over time, test and Teacher Assessment levels have become very close at KS1.
- There are important questions to answer about the use of the various forms of assessment at key stage 1 and value-added calculations carried out later using these data.

In relation to the sample of LEAs and schools and the methodology:

- The sample of LEAs and schools appear to be representative.
- The wide range of methods and sources of evidence used in the evaluation, and the numbers involved, should ensure that the findings have validity and some depth.
- On this basis, the presentation of the results that follow in the subsequent chapters would seem to be both valid and generalisable.

2. The ‘robustness’ of the trial arrangements

Assessment at the end of key stage 1 is significant in policy terms. Within the wide spectrum of national data collection on educational performance, it provides a pivotal point for calculations of value-added. It must be, and be seen to be, ‘robust’. But what is the meaning of this word in this context?

The assessment arrangements of the trial replaced a separate reporting of Teacher Assessment on the one hand and test and task results on the other, with reporting just one set of results - Teacher Assessments - that have in all cases been informed by test and task outcomes.

Beyond reporting to parents and being used to monitor the progress of individual children within and across schools, KS1 assessment results are used to track trends of performance at the school level, through ‘PANDA’ analysis and ‘value-added’ measures. It is important, therefore, that the changes to assessment arrangements do not compromise the capacity of the assessments to fulfil these functions. That is the meaning of ‘robustness’ in this context.

Robustness in tasks and tests

The tests and tasks at KS1 constitute a complex set of related assessments in English and mathematics. In reading and mathematics the tests and tasks each operate only at certain levels, but in writing the task covers all levels, as does the spelling test. In the existing (pre-trial) system there are strict rules about which tests and tasks are to be administered to which children, when and in which order. In the trial there were fewer requirements of this kind, and greater choice about which test or task was used (so long as every child did undertake at least one in each area).

In both systems, the tasks may be administered by the teacher at any point over a long time period (January to June each year). The tests are not time-limited. In the trial arrangements schools were able to choose between the 2003 and 2004 tests. The 2004 tests could be administered over a given time period in May, but those for 2003 could be used earlier. The tests and tasks are marked by the teachers themselves, not by an outside agency. In the existing system LEAs were required to provide training for marking, and then to audit the process, especially in relation to the tests. In the trial schools, the emphasis of moderation was on the whole process of arriving at an accurate assessment judgement, not specifically on the tests. In fact, QCA’s advice was to not ‘audit’ tests and their marking.

Against this background, ‘robust’ tests and tasks are those:

- which cover as much as possible of the KS1 curriculum in a balanced way and at an appropriate level of difficulty, bearing in mind also the provisions of the National Literacy Strategy, the National Numeracy Strategy, and all subsequent initiatives related to these;
- which have been carefully developed, trialled and pre-tested;
- where pupils are comfortable and familiar with the test-taking/task assessment situation;
- which are administered in a fair way – including, for tasks, issues in group administration;
- which are marked fairly, consistently and comparably.

Robustness in Teacher Assessment

The general assumption of Teacher Assessment is that teachers will assess their pupils in a formative way throughout Year 2 and that they will be retaining evidence of pupil performance in a systematic way, through the record-keeping systems in use in the school. These assessment records, together with the evidence available to teachers in the form of children's work, for example in their workbooks, are then drawn on in coming to a judgement about attainment levels towards the end of the year. In the existing arrangements, teachers report their Teacher Assessments separately from the results of tasks and tests, and teachers can come to a judgement about levels of performance without considering the outcomes from tests and tasks, and report it, even if the teacher judgement of the level of performance differed from the test and task outcome. In the trial, only one assessment level for each child in each subject was reported, so the judgements of Teacher Assessment had to take the evidence from tests and tasks into account, to provide a rounded, summative judgement at the end of the year.

Against this background, 'robust' Teacher Assessment will be an assessment:

- which covers as much as possible of the KS1 curriculum and Attainment Targets in a balanced way, bearing in mind also the provisions of the National Literacy Strategy, the National Numeracy Strategy, and all subsequent initiatives related to these;
- which takes into account all the appropriate evidence, in a fair and considered way;
- which is appropriately moderated within schools and across schools to ensure that consistent standards are being applied;
- which arrives at defensible judgements, about which there is widespread confidence.

It is widely recognised that schools need support in meeting some of these requirements, and the role of LEAs in providing appropriate training and moderation is very important. It is also an important feature of robustness that the results are accurately and appropriately reported, both to the national data collection agency and to governors, pupils and parents.

These considerations indicate that there is scope for validity, reliability, and management requirements **not** to be met, and hence for the robustness of the assessments to be compromised, both in the previous KS1 assessment context and in the new one.

The evaluation addressed the issue of robustness in different ways, which will be reported in subsequent sections of this chapter.

2.1 A comparative analysis of the data in 2003 and 2004 of Year 2 pupils in the schools in the evaluation sample

Comparative analysis of the patterns of performance
The effect of LEA on performance
The effect of gender on performance
The effect of age on performance
School-type, value-added and PANDA
Variation between task/test results and Teacher Assessment

2.2 An evaluation of the training and moderation provided for schools in the LEAs of the evaluation sample in 2004

Training
Moderation

2.3 Other evidence to evaluate the robustness of the data from the trial system

Evaluation schools data and the DfES data – a comparison
The confidence of headteachers and LEA assessment coordinators

2.4 Key issues and findings

2.1 A comparative analysis of the data in 2003 and 2004 of Year 2 pupils in the schools in the evaluation sample

Data from 2003 and 2004 on the performance of the Year 2 children in the 107 schools across 10 LEAs in the evaluation sample have been analysed in a variety of ways, and will be reported in the sections following.

2.1.1 Comparative analysis of the patterns of performance

2.1.2 The effect of LEA on performance

2.1.3 The effect of gender on performance

2.1.4 The effect of age on performance

2.1.5 School-type, value-added and PANDA

2.1.6 Variation between task/test results and Teacher Assessment: Cohen's Kappa and teacher/headteacher perceptions

The 2003 data used in these sections are extracted from the database supplied by DfES, reflecting the returns made by the schools to LEA. There are data on Teacher Assessments for 4265 children and on tasks and tests for 4268 children in the evaluation schools.

The 2004 data used in these sections comes directly from the schools. As part of the trial arrangements schools are not required to send test and task results to the LEA/DfES. In order to obtain this information the evaluation team dealt with the evaluation schools directly, and the following analyses are based on these data supplied by schools. There is Teacher Assessment data for approximately 4100 children (4151 in speaking and listening; 4100 in reading; 4087 in writing and 4119 in mathematics) and task and test data for approximately 3050 children (4268 in reading; 3066 in writing and 3053 in mathematics).

Note

When the data were requested from schools, they were asked for reading test and task data separately, to conform to last year's data. However, when the data were received it was clear that the two had been reported in one column, sometimes under task and sometimes under test, as both contain levels of pupil attainment not associated with this type of activity, i.e. level 3 for task, and level 1 and W for test. For both 2003 and 2004 where a task and a test result have been given the data have been combined by taking the higher of the two levels given.

2.1.1 Comparative analysis of the patterns of performance

Two measures are used to compare performance in 2003 with that in 2004. One is the level awarded to each child, and the proportions of the sample obtaining each level, to see whether proportions have changed. The other is a mean value, obtained by applying the scaling system shown in Table 1.9 to each award of a level, finding the mean for all the children in the sample, to see whether there has been a change.

Comparing the mean value of levels awarded

Table 2.1 uses the mean values to compare Teacher Assessment and test and task results for the years 2003 and 2004. Statistical significance is indicated by asterisks in the conventional way:

- * at the 5% level
- ** at the 1% level
- *** at the 0.1% level

Table 2.1 Comparative mean performance data for 2003 and 2004, in Teacher Assessment and Tasks and Tests (Evaluation schools)

Topic	Teacher Assessment		Task / Test	
	2003	2004	2003	2004
Speaking & Listening	6.457	6.495		
Reading	6.518	6.650**	6.663	6.589*
Writing	6.002	6.014	6.040	6.012
Maths	6.584	6.650*	6.816	6.712**

The standard deviation for each of the means is in the region of 2, and are all comparable year on year (to within 5%).

Examining Table 2.1 reveals a number of features. The mean Teacher Assessment in 2004 is higher than that in 2003 in all subjects, including speaking and listening, which was not subject to any change in arrangements (as there are no tests or tasks, and was just Teacher Assessment in both years). As indicated by the asterisks, the difference is significant in reading and maths. In test and tasks, however, the mean performance is lower in 2004 compared to 2003, and is significantly lower in mathematics and reading. There is also a tendency in both years for the Teacher Assessment to be lower than the task and test result, although in reading and writing in 2004 the mean Teacher Assessment was slightly higher. It should be noted in all of this that even when the differences are statistically significant, they are small in relation to actual performance and to the standard deviation of the sample.

Comparing the proportions of levels awarded

Table 2.2 and Figures 2.1 – 2.4 give the percentage of pupils attaining a particular level or sub-level in each subject.

Table 2.2 Percentage of pupil attaining levels in the assessed subjects in 2003 and 2004

Assessment		Percentage at each level							N
		W	1	2C	2/2B	2A	3	4	
Speaking & Listening	2003 TA	1.4	9.9		61.3		27.3	0.1	4265
	2004 TA	1.2	9.2		62.2		27.4		4151
Reading	2003 TA	2.4	11.4		53.6		32.4	0.2	4265
	2004 TA	2.8	9.4	13.8	57.8/21.3	22.7	30.0		4100
	2003 Test/task	2.4	11.6	8.2	56.3/17.9	29.2	30.7		4268
	2004 Test/task	2.2	8.9	13.7	60.2/21.7	24.8	28.6		3053
Writing	2003 TA	4.0	11.7		65.9		18.3		4265
	2004 TA	4.4	9.8	20.7	68.6/26.4	21.5	17.3		4087
	2003 Test/task	4.4	12.2	18.0	63.7/24.0	21.7	19.6		4268
	2004 Test/task	4.3	10.5	20.7	67.3/25.4	21.2	17.9		3066
Maths	2003 TA	1.3	7.8		61.5		29.3		4265
	2004 TA	1.9	5.9	16.6	64.3/22.7	25.0	27.9		4119
	2003 Test/task	1.5	5.6	16.7	60.3/19.1	24.5	32.6		4268
	2004 Test/task	1.9	5.6	16.2	62.5/21.9	24.4	29.9		3053

Figure 2.1. Pupil attainment in Speaking and Listening

Pupil Attainment in Speaking and Listening, Reported Assessment

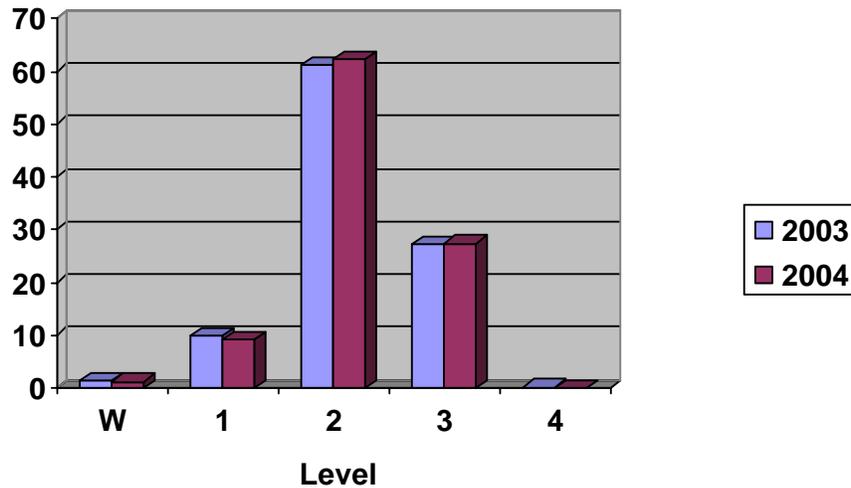


Figure 2.2. Pupil attainment in Reading

Pupil Attainment in Reading, Reported Assessment

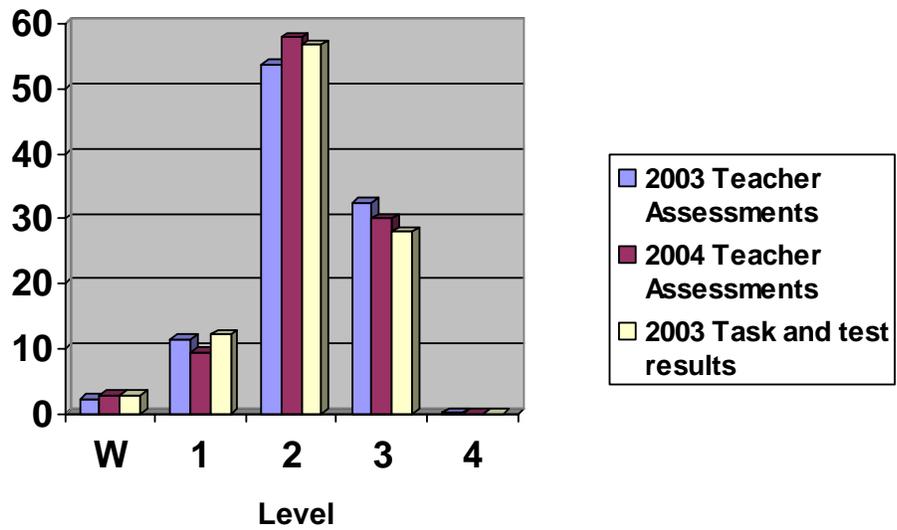


Figure 2.3. Pupil attainment in Writing

Pupil Attainment in Writing: Reported Assessment

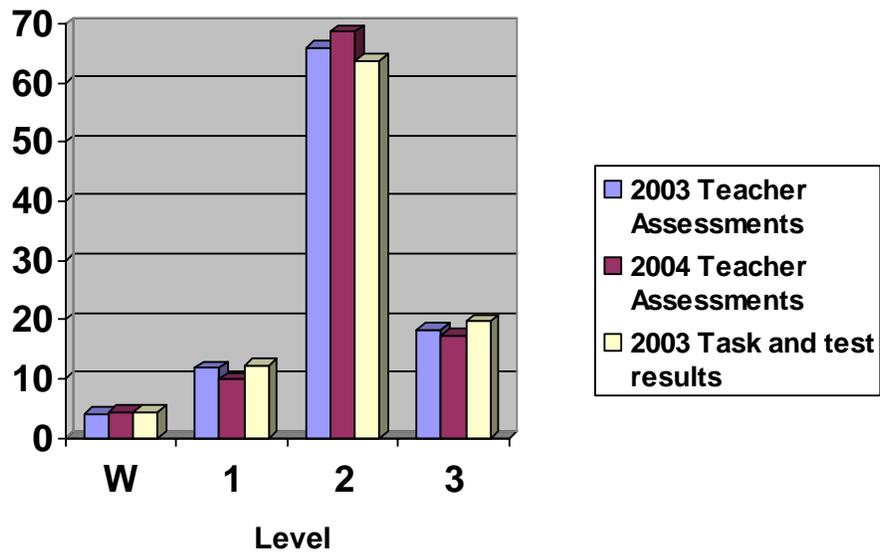


Figure 2.4. Pupil attainment in Maths

Pupil Attainment in Maths: Reported Assessment

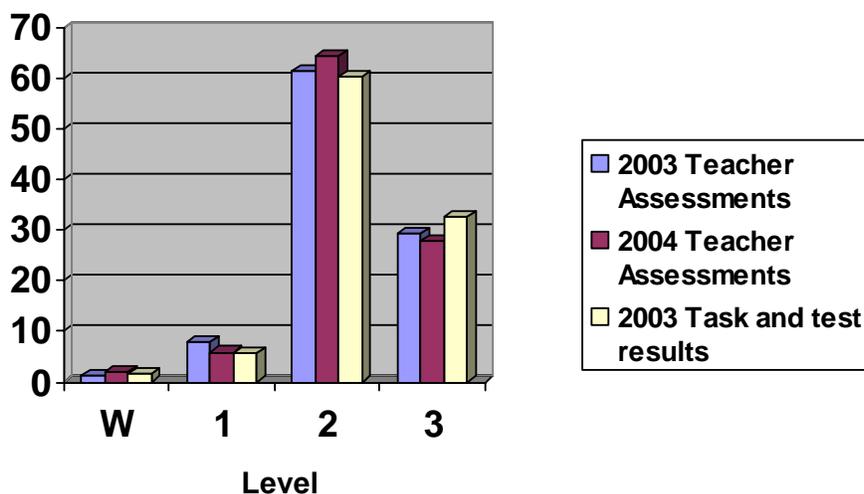


Table 2.1 and Figures 2.1-2.4 show that, in all subjects, the number of pupils attaining level 2 in 2004 is greater than was reported in 2003 for either Teacher Assessment or test and task result. All of these differences are significant at the 0.1% level except in speaking and listening, where the difference is not statistically significant. There are also fewer children reported as level 1 in all subjects in 2004, except compared to test and task in mathematics in 2003, and fewer children reported as level 3 in 2004, except compared to test and task in reading in 2003.

It is possible that this is an effect of the change to assessment arrangements, particularly as the differences are not so marked in speaking and listening, which was not directly affected by the new arrangements. However, it does not seem to arise solely from teachers' discretion in Teacher Assessment, since the 2004 test and task results for all three subjects also show the trend of more children at level 2 and fewer at levels 1 and / or 3, with the differences in reading and writing significant at the 0.1% level, and those for mathematics at the 1% level.

Nevertheless, it may legitimately be asked what the change represents in relation to the robustness of the results. Is it an example of the statistical phenomenon of *end-aversion bias* (Streiner and Norman, 1995) where the splitting of level 2 into three sub-levels has made levels 3 and 1 seem more extreme, and were therefore avoided by teachers - at the expense of accuracy in the assessments? Or might it be the result of greater accuracy in Teacher Assessments in the trial system? There was certainly a general feeling expressed by many stakeholders that the assessments, arrived at as a result of the new assessment arrangements, are more accurate than test and task results alone.

The question cannot be resolved by simple 'objective' means. Any view is likely to be affected by opinions about the qualities of test and task assessment on the one hand and teacher judgement on the other, informed by indicative (rather than conclusive) evidence. For example, the smaller proportion of children in the evaluation schools achieving level 3 in the tests and tasks in 2004 compared to 2003 could be taken to reflect a change in the abilities of the cohort from last year. On the other hand it might be argued that a reduced emphasis on testing might have led to less preparation for the tests, and therefore that some children may have done less well than they would have done if prepared more intensely. This then begs the

Evaluation of the trial assessment arrangements for key stage 1

question of whether such children would have been ‘really’ performing at that level, or would have had their performance boosted to a level that does not reflect their true competence.

When comparing the reported Teacher Assessments in 2004 with test outcomes, there is again a slight tendency to more Teacher Assessment at level 2. One possible justifying argument for fewer level 1 assessments may be that children working at level 1 are more inclined to respond badly to test contexts, and so are prone to under-perform in that one-off situation, compared to what they show they can do over time in class. The equivalent argument ‘explaining’ a lower incidence of level 3 assessment is that it is possible to garner enough marks to be awarded level 3 on a test through success on work that is at level 2. Test thresholds presume a level of error in performance, and so a well motivated, well focused, well prepared child who is generally working at level 2A can attain level 3 in a one-off situation.

The opposing arguments say that teachers in class sometimes give more help than they realise to children operating at level 1, and inflate their judgements of the children’s work as a result of sympathy, and that teachers operating in a whole class context with a set curriculum sometimes constrain the work of children who are capable of operating at level 3, so that it is not possible for the children to show the full range of level 3 work in class, as they are not given the opportunity to do it.

The change to Teacher Assessments in 2004 compared to 2003 (more at level 2 and fewer at levels 1 and 3) may be a cohort effect, or reflect greater accuracy as a result of more focused attention on Teacher Assessment, or be because in 2003, under the regime of separate reporting, there were reasons in many individual cases for emphasising the difference between Teacher Assessment and test result (in order to give a more accurate balanced impression when both were read together), whereas in 2004 the Teacher Assessment was the only reported assessment, and the ‘balancing’ occurred before reporting rather than in reporting.

The tensions between the different points of view cannot be fully resolved by evidence, and remain possibilities in each particular case. As such, the value of having checks and balances in the system is highlighted. The importance of tasks and tests as a standard national calibration should not blind us to the uncertainties involved in applying the measure in individual cases. Equally the importance of teachers’ professional judgement as the most valid source of insights about an individual learner should not blind us to the possibilities of sympathetic bias and relativism. The new assessment arrangements offer the potential to have the best of both worlds, enabling teachers to act on doubts about the validity of specific test results, but also with test and task outcomes (and moderation) available as checks that keep the professionals questioning.

Considering all the discussion above, it seems a well founded (but not certain) proposition that the change to the population profile of assessment outcomes is the result of the assessments in the trial being more accurate.

Summary of comparative performance

- The mean of Teacher Assessments for the Year 2 children in the schools of the evaluation sample are higher in 2004 than in 2003, but the mean test and task results in 2004 are lower than test / task results in 2003.
- In the evaluation schools more children, and in all subjects, more children were awarded a Teacher Assessment at level 2 than were awarded level 2 in either Teacher Assessments or test and task result in 2003. Generally there were also fewer level 1s and 3s awarded in 2004 than in 2003.

2.1.2 The effect of LEA on performance

The trial system was implemented in 34 different LEAs around the country, ten of which took part in the evaluation. This section examines the data from evaluation schools to consider whether there were differential changes in performance across evaluation LEAs, which might represent an effect of what different LEAs did in training, moderation and other support for schools. If variations were to be found, this would raise questions about the robustness of the new arrangements.

Figures 2.5 and 2.6 compare the mean values of the Teacher Assessments in evaluation schools in different subjects in 2003 and 2004, broken down by LEA.

Fig 2.5 Mean values of the Teacher Assessments in the evaluation schools of the ten evaluation LEAs for 2003

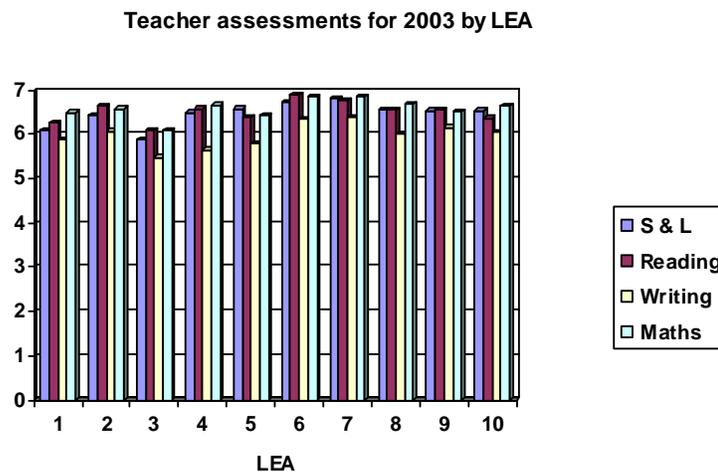
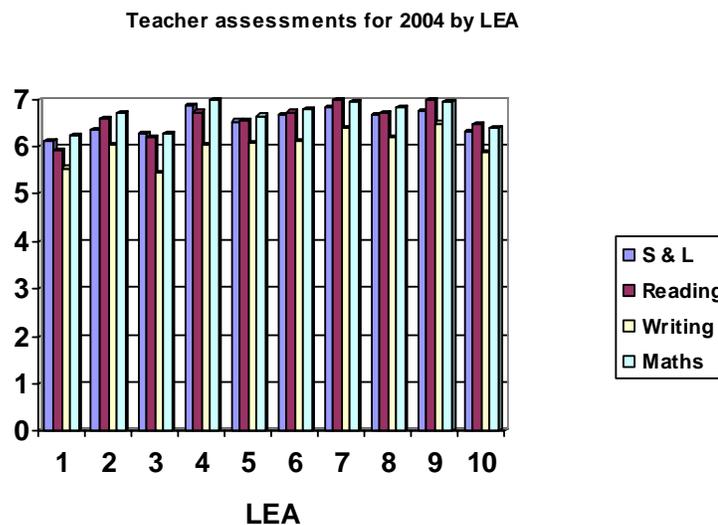


Fig 2.6 Mean values of the Teacher Assessments in the evaluation schools of the ten evaluation LEAs for 2004



Evaluation of the trial assessment arrangements for key stage 1

It is clear from Figures 2.5 and 2.6 that the mean performance levels of the evaluation schools in different LEAs can vary by as much as a value of 1 on the scale, which represents an average difference of one sub-level (e.g. 2C to 2B or 2B to 2A).

Did the trial bring about different changes in some LEAs compared to others? Table 2.3 compares the mean values of levels given to pupils in the evaluation schools in 2003 and 2004, broken down into different LEAs.

Table 2.3 Mean performance levels for 2003 and 2004 by LEA for evaluation schools

LEA	Speaking and listening		Reading		Writing			Maths		
	2003 TA	2004 TA	2003 TA	2004 TA	2003 TA	2004 TA	2003 Task	2003 TA	2004 TA	2003 Test/task
1	6.10	6.12	6.27	5.92	5.87	5.55	5.78	6.50	6.23	6.69
2	6.42	6.34	6.64	6.60	6.08	6.04	6.16	6.58	6.70	6.90
3	5.89	6.27	6.09	6.19	5.48	5.46	5.46	6.09	6.28	6.43
4	6.50	6.88	6.58	6.73	5.65	6.05	5.64	6.67	7.00	6.85
5	6.58	6.53	6.40	6.54	5.79	6.07	5.87	6.45	6.65	6.73
6	6.73	6.67	6.90	6.73	6.36	6.13	6.31	6.86	6.78	6.88
7	6.82	6.82	6.76	6.97	6.40	6.38	6.62	6.86	6.96	7.05
8	6.57	6.66	6.56	6.70	6.01	6.20	6.10	6.70	6.83	6.93
9	6.54	6.74	6.57	7.00	6.16	6.49	6.11	6.52	6.95	6.84
10	6.54	6.33	6.37	6.46	6.07	5.88	6.16	6.66	6.38	6.85

In speaking and listening (which was not directly affected by the trial) there is slight variation up or down in the Teacher Assessments from 2003, by up to 0.38, which is significant at the .01 level (LEAs 3 and 4).

In reading there is again variation in the mean Teacher Assessment score, both up and down, and by as much as 0.43 (in LEA 9), significant at the .001 level.

In writing, the difference in mean score can be as much as 0.4 when 2003 Teacher Assessment is compared to 2004 Teacher Assessment (significant at the .01 level), and as much as 0.41 when 2003 task score is compared to mean 2004 Teacher Assessment (significant at the .01 level). However these cases both apply to LEA 4, where the writing task score in 2004 for the evaluations schools was particularly high, as can be seen in Table 2.4 (with a statistically significant difference at the .001 level compared to the 2003 writing task score). Since the 2004 Teacher Assessment takes the writing task into account in arriving at a final judgement, the effect seems to arise from differences in actual performance, rather than from anomalies in judgement.

In mathematics, the maximum difference in mean score between 2003 Teacher Assessment and 2004 Teacher Assessment is 0.43 (LEA 9, significant at the .001 level) and that between 2003 test / task score and 2004 Teacher Assessment is 0.46 (LEA 1, significant at the .01 level). However, the variations in mathematics may again be seen to reflect different performance in the mathematics tests and tasks, with the mean 2004 Teacher Assessment being slightly lower than or the same as the mean mathematics test and task score in all LEAs. This can be seen from the results in Table 2.4. The variation from 2003 seems again therefore to arise from differences in performance of the children in the evaluation school samples.

Table 2.4 Mean performance in writing and mathematics tasks / tests for evaluation schools, by LEA

LEA	Writing Task 2004	Maths Test/task 2004
1	5.53	6.48
2	5.97	6.80
3	5.36	6.30
4	6.39	7.09
5	5.93	6.67
6	6.24	6.78
7	6.35	7.00
8	6.26	6.83
9	6.36	6.99
10	5.76	6.38

There is a different year on year change in the mean level scores of the children in evaluation schools from some LEAs compared to others. From the data available within the study it is not possible to determine whether this is a systematic variation due to differences in LEA activities, or is a normal level of change due to random factors. However, the following points suggest:

- the differences are relatively small in relation to levels;
- the differences also apply to an extent to speaking and listening, which was not affected directly by the new arrangements in the trial;
- the differences seem to reflect closely the pattern of test and task results.

2.1.3 The effect of gender on performance

The purpose of this section is to examine whether the trial seems to have had effects on performance that can be differentiated by gender. It could be supposed, for example, that girls might benefit from arrangements that emphasised classwork compared to tests.

Table 2.5 shows the mean level scores of children in the evaluation school split between boys and girls. The data have been corrected for age differences (in case girls in the sample happened to be younger or older on average than the boys), with a common age of 7 years 3 months at first of May taken as representative.

Table 2.5 The mean performance of boys and girls attending evaluation schools

	2003			2004		
	Boys	Girls	Diff	Boys	Girls	Diff
Speak /listen TA	6.293	6.653	-0.36***	6.326	6.761	-0.435***
Reading TA	6.252	6.791	-0.539***	6.294	6.889	-0.595***
Writing TA	5.692	6.298	-0.606***	5.639	6.430	-0.791***
Writing task	5.666	6.393	-0.727***	5.609	6.446	-0.837***
Maths TA	6.642	6.535	+0.107	6.696	6.664	+0.034
Maths Test / task	6.848	6.782	+0.066	6.766	6.707	+0.059

The 2003 data show that girls already did significantly outperform boys on literacy topics, and that boys did already out perform girls in mathematics, although not significantly. In 2004 the picture is maintained, with girls slightly increasing the difference in English (but not

Evaluation of the trial assessment arrangements for key stage 1

significantly, and including speaking and listening, which was not directly affected by the trial) but with boys' advantage in mathematics slightly eroded. Both of these differences arise from a small, non-significant, but consistent increase in performance across all areas for girls, but not for boys. As this difference applies to task and test results, as well as to Teacher Assessments, either a genuine trend or a sample effect is suggested, rather than any effect arising from the trial.

Interaction between LEA and gender

A further analysis was undertaken to explore the possible interaction between effects of gender and of LEA - is the difference in performance between boys and girls consistent across all LEAs? Table 2.6 shows some selected results of these calculations.

Table 2.6 The interaction between gender and LEA for different assessments in 2004 (Evaluation schools)

Subject	Interaction significance value
Speaking & Listening Teacher Assessment	0.818
Reading Teacher Assessment	0.104
Writing Teacher Assessment	0.114
Writing task	0.055
Maths Teacher Assessment	0.792
Maths task and test	0.990

There does not seem to be a significant interaction between gender and LEA.

2.1.4 The effect of age on performance

It is widely felt that in early schooling, the summer born children (May to August) can be at a disadvantage compared to autumn born children (September to December) because at any point they have had less time in school, and are generally less 'mature' and so less able to take advantage of the provision being made for them.

This section examines the possibility that age has a differential effect on the assessment processes in the new arrangements – whether the form of assessment is particularly advantageous or otherwise to children who are relatively young or relatively old within the cohort. Table 2.7 looks at age effects for different assessments for the years 2003 and 2004. In these calculations the age of children has been treated as a continuous variable although in practice any age effect will be affected by school policies on school entry, such as that children will only enrol at the start of a term.

Table 2.7 The effects of age on different assessments (evaluation schools)
(*Mean age 7 years 3 months at 1st May*)

Assessment	2003		2004	
	The mean score differential for one year age difference	The percentage of performance variation attributable to age	The mean score differential for one year age difference	The percentage of performance variation attributable to age
Speak/listen TA	0.900	2.0***	0.959	2.4***
Reading TA	1.106	2.4***	1.192	2.9***
Writing TA	0.919	1.9***	1.181	3.1***
Maths TA	1.044	2.8***	1.309	4.3***
Writing task	1.199	2.8***	1.422	4.3***
Maths task/test	1.239	3.7***	1.494	5.5***

The results shown in Table 2.7 indicate that the age factor is significant (at the 0.1% level) in both 2003 and 2004, being the equivalent of approximately one sub-level of performance (e.g. 2C to 2B) for a full year of difference in age. Age accounts for up to almost 6% of the variance in the performance of the children.

There are also variations in comparative values. In all cases the age effect is greater in magnitude and influence for the task and test compared to the Teacher Assessments, and also greater in magnitude and influence for 2004 compared to 2003.

It is possible that the enhanced age effect in 2004 arises because age has a greater effect on the performance of children in tests than it does in teacher assessment based on classwork – perhaps because the efforts of the teacher in supporting the less ‘mature’ individuals to show their true potential in classwork is not available in tests. The requirement to take task and test into account in 2004 in arriving at the final Teacher Assessment may then explain why there appears to be a greater effect on Teacher Assessment in 2004 compared to 2003. In this respect the new arrangements may bring a softening of the age effect on the assessments used for PANDA and ‘value-added’, compared to the use of test and task results alone.

Summary of effect of LEA, gender and age on performance

- The performance across LEAs varies.
- There is variation in the differential Teacher Assessment from 2003 to 2004 across LEAs. However these differences also apply to speaking and listening and to test and tasks.
- The trial does not appear to have affected gender differentials. The pattern in 2004 is similar to 2003 and both test/task and Teacher Assessment patterns in 2004 are similar to each other.
- There is no significant interaction between LEA and gender.
- Age does seem to have some effect, although there is some uncertainty as to why this may be.

2.1.5 School type, value-added and PANDA

One of the important features of robustness is the ‘fitness for purpose’ of the assessment outcomes in relation to ‘value-added’ and PANDA. One of the key concerns in this area is the possible effect due to type of school. In Infant and First schools, the judgements about success are based more on KS1 results than is the case in Primary schools, for whom the ‘value-added’ progress from Year 3 to Year 6, comparing the KS1 results with those at KS2 is

Evaluation of the trial assessment arrangements for key stage 1

perhaps more important. It reflects well on an Infant school if many children achieve high levels at KS1 (although that might then impact negatively on the Junior school to which the children transfer), but in Primary schools it can be better overall for a school if fewer children achieve high levels at the end of key stage 1. There has been some concern expressed in the questionnaires and interviews that under the pressure that arises from this situation, Infant and First schools may tend towards being 'generous' / 'optimistic' in their Teacher Assessments, for example when a child is on the borderline between level 2A and level 3, whereas Primary schools may not.

A comparative analysis was undertaken using the evaluation schools of the sample, comparing the results from Infant and First schools to those from Primary schools. The results of this analysis are shown in Tables 2.8 to 2.12

Table 2.8 Showing the percentage of reported level 3 Teacher Assessment in each subject, by type of school (2004, evaluation schools)

	Primary schools	Infant and First schools
Speaking and listening	26.9 (n=2438)	28.0 (n=1721)
Reading	27.8 (n=2407)	32.0 (n=1719)
Writing	15.8 (n=2407)	18.7 (n=1719)
Maths	25.9 (n=2427)	30.4 (n=1718)

Table 2.8 shows that the proportion of children being given a Teacher Assessment of 3 is higher in Infant and First schools than it is in Primary schools. This may be a true reflection of their attainment and represent a sample effect, or it could arise from different choices made in different kinds of schools. It will be of value to examine the Teacher Assessments in relation to the test results. Table 2.9 shows the test and task results for the schools.

Table 2.9 The percentage of reported level 3 test and task results in each subject, by type of school (2004, evaluation schools)

	Primary schools	Infant and First schools
Reading	27.6 (n=1759)	29.8 (n=1293)
Writing	16.8 (n=1791)	19.2 (n=1297)
Maths	27.3 (n=1759)	33.1 (n=1304)

In all cases the Infant and First schools had higher test results than the Primary schools, and in all subjects and in both kinds of school the Teacher Assessments are slightly below the test results in terms of numbers of level 3s awarded.

The nature of decisions in borderline cases can also be examined by comparing test and task result with Teacher Assessments. In Table 2.10 the numbers for reading are reported.

Table 2.10 Variation between test/task result and Teacher Assessment in reading, by type of school, in evaluation schools

		Reading Test & Task							Total
		W	1	2C	2B	2A	3		
Primary schools	Reading TA	W	38	0	0	0	0	0	38
		1	7	147	13	2	1	0	170
		2C	0	8	194	17	3	0	222
		2B	0	2	19	337	33	1	392
		2A	0	0	1	27	365	22	415
		3	0	1	0	0	40	451	492
	Total	45	158	227	383	442	474	1729	
Infant and first schools	Reading TA	W	22	0	1	0	0	0	23
		1	1	105	6	0	0	1	113
		2C	0	6	172	11	0	2	191
		2B	0	0	9	255	10	0	274
		2A	0	2	1	10	263	6	282
		3	0	0	1	0	32	376	409
	Total	23	113	190	276	305	385	1292	

The number of children being given level 3 for reading as a Teacher Assessment whose test and task result was lower than 3 is 41 (2.4%) in Primary schools and 33 (2.6%) in Infant and First schools.

Table 2.11 shows the equivalent results for writing.

Table 2.11 The number of children whose task result differs or agrees with their Teacher Assessment in writing, by type of school in evaluation schools

		Writing Task							Total
		W	1	2C	2B	2A	3		
Primary schools	Writing TA	W	77	0	0	0	0	0	78
		1	7	156	11	0	1	0	175
		2C	0	34	315	12	1	1	363
		2B	0	2	44	408	27	3	484
		2A	0	2	1	25	328	18	374
		3	0	0	1	1	12	272	286
	Total	84	195	372	446	369	294	1760	
Infant and first schools	Writing TA	W	44	3	0	0	0	0	47
		1	3	125	4	2	0	1	135
		2C	0	18	234	19	3	0	274
		2B	0	1	19	314	30	1	365
		2A	0	0	0	11	234	26	271
		3	0	0	0	0	8	221	229
	Total	47	147	257	322	275	249	1297	

Here the number of children being awarded a Teacher Assessment of 3 without a test / task result at level 3 is 14 (0.8%) for Primary schools and eight (0.6%) for Infant and First schools.

Finally, Table 2.12 gives the results for mathematics.

Table 2.12 The number of children whose test/task result differs or agrees with their Teacher Assessment in maths, by type of school in evaluation schools

		Maths Test & Task							Total
		W	1	2C	2B	2A	3		
Primary schools	Maths TA	W	36	4	0	1	0	0	41
		1	8	100	8	0	1	0	117
		2C	0	11	241	24	3	1	280
		2B	0	0	24	364	43	2	433
		2C	0	0	2	13	377	29	421
		3	0	0	0	0	13	444	457
	Total	44	115	275	402	437	475	1749	
Infant and first schools	Maths TA	W	11	0	0	0	0	0	11
		1	4	53	9	2	0	0	68
		2C	0	7	189	22	1	0	219
		2B	0	0	19	226	32	2	279
		2A	0	0	2	16	278	36	332
		3	0	0	0	0	1	394	395
	Total	15	60	219	266	312	432	1304	

Here 13 (0.7%) children in Primary schools were awarded a Teacher Assessment level of 3 when their test result was below 3, but in Infant and First schools it was only one (less than 0.1%).

In order to investigate this issue further the 2003 data have been examined, to assess whether the higher proportion of level 3s in Infant and First schools was apparent prior to the introduction of the new trial arrangements.

Table 2.13 Showing the percentage of reported level 3 Teacher Assessment in each subject, by type of school (2003)

	Primary schools (n=2327)	Infant and First schools (n=1729)
Speaking and listening	24.5	31.3
Reading	30.9	34.5
Writing	15.7	22.2
Maths	26.6	33.4

In 2003 the children in the Infant and First schools among the sample also had higher reported Teacher Assessment than the children in Primary schools in the sample. Again it is of value to examine the test results as well, and these are shown in Table 2.14.

Table 2.14 The percentage of reported level 3 test and task results in each subject, by type of school (2003)

	Primary schools (n=2328)	Infant and First schools (n=1731)
Reading	29.8	31.4
Writing	16.8	23.1
Maths	31.0	35.1

Table 2.14 shows that the test and task results were also higher for Infant and First schools.

Evaluation of the trial assessment arrangements for key stage 1

In short, concern about Infant and First schools inflating their Teacher Assessments appears to be unfounded. If there are any incidences of schools inflating the performance of their pupils for PANDA purposes this has not occurred as a result of the trial arrangements.

When this issue was raised in interviews or in teacher focus groups, there was an acknowledgement of the nature of the pressure. One teacher for example said:

you do get a bit of pressure from both side, because you get pressure from the Juniors saying, "We don't want all these threes, because then we can't progress" and then your head says to you, "Well ... what about our results?" I mean the league tables [and] when the patch inspector comes in ... setting the school targets based on your result.
(teacher)

However, there is also clarity about what has to occur, and why:

I don't think that [the target setting agenda] means you'd report unfairly on your children.

(teacher)

You would be making a rod for your own back.

(teacher)

You're not being fair to your colleagues.

(teacher)

There is no point in saying they are a 2A if they are a 2B because when they go up into year 3 that's not helping that child.

(teacher)

There was also a widespread view that the new arrangements, while perhaps giving more opportunity to interpret results, also give more evidence to resist pressure to interpret them in a certain way:

I think that that's where the teacher assessment will actually give us more strength to say that this is a fair and accurate far more fair and accurate than probably some of the test results were.

(teacher)

Summary of school type, value-added and PANDA

- In 2004 the Infant and First schools in the sample report a higher proportion of their children achieving level 3 in all subjects than the Primary schools.
- The test and task results of the children in the Infant and First schools in the sample also had more level 3s than those in Primary schools, in all subjects, and in all cases the Teacher Assessment proportion was lower than the test and task result proportion.
- The numbers of children being awarded level 3 who did not attain level 3 in the test / task are very low in all subjects for both kinds of school, but are lower in Infant and First schools than in Primary schools.
- In 2003 there were similar differences between the proportions of level 3s in the Infant/First schools and the Primary schools in the sample.
- Teachers are aware of the conflicting pressures on them, but do not allow this to affect their professional judgements.

2.1.6 Variation between test / task results and Teacher Assessment

Cohen's Kappa is a tool used to assess the extent to which two measures agree. It is used here to quantify the agreement between test/task results and Teacher Assessments in 2003 and 2004. If tests and tasks and Teacher Assessments are separately thought to be robust, one should expect high levels of agreement between the two measures of children's achievement.

There are, of course, good reasons why the two measures might not be the same (see the review of the literature in Chapter 1), but the application of this analysis provides a perspective on the robustness of the trial assessment arrangements, by comparing the results in 2004 to those of 2003, when the 'old' arrangements applied. Of course, Teacher Assessment in 2003 and the Teacher Assessment in 2004 are not directly comparable, since in 2003 Teacher Assessment was in theory an independent measure (even though in practice many teachers did base their Teacher Assessments on the task / test result, as reported in Chapter 1) whereas in 2004 Teacher Assessment is required to be informed by the task/test result and therefore is not independent. Therefore in 2004 one would expect a higher degree of agreement between Teacher Assessment and test and task result. A reduced agreement would raise questions about robustness in the new system.

The Kappa statistic essentially compares the amount of agreement actually found (the observed agreement) with the expected agreement. The 'expected' agreement is the amount of agreement that would occur if the test/task and Teacher Assessment were completely independent.

If the observed agreement is equal to the 'expected' agreement (i.e. the amount of agreement one would expect by chance) then Kappa takes the value of 0 and if there is 100% agreement then Kappa takes the value of 1. A suggested set of benchmarks for the Kappa statistic (Landis and Koch, 1977) are given below, though there is always an element of subjectivity in the judgement.

0.00 - poor	0.41-0.60 - moderate
0.01-0.20 - slight	0.61-0.80 - substantial
0.21-0.40 - fair	0.81-1.00 - almost perfect

It is suggested in literature (e.g. Sheskin, 2004) that where categories of data are sequential, then weighting should be used to indicate levels of agreement. Weightings have not been used in this investigation for the following reasons:

- it would invalidate comparisons with existing studies, which tend to use the simple model;
- the small number of categories, coupled with their width, make disagreement in anything other than adjacent categories unlikely.

In order to make the comparison, each set of results has to have exactly the same categories. For the 2003 data the Teacher Assessment data are not given at a sub-level for the Level 2 children, whereas the task and test data are. Therefore the levels 2A, 2B and 2C have been collapsed to level 2 and the levels (for example level 4) which are not represented in both kinds of results have been excluded from the analysis. The 2004 data are given as sub-levels for test/task and Teacher Assessment and is therefore analysed on this basis, but is also shown in 'collapsed' form for easier comparison.

Kappa calculations: the 2003 data

First there is an examination of the variation between test/task and Teacher Assessment under the current system, using 2003 data for the children in all the schools in the trial, for the children from all the schools in the evaluation LEAs, and for children in the evaluation schools. For the schools in the evaluation the data are presented in more detail, in tables showing where the agreement and disagreement occurs.

1) Mathematics

Table 2.15 shows the comparative data of task/test results and the Teacher Assessments, on which Cohen's Kappa is calculated.

Table 2.15 Comparison of Teacher Assessment and test/task for schools in evaluation sample, 2003 data, in mathematics

		Mathematics task/test				Total
		W	1	2	3	
Maths TA	W	42	12	2	0	56
	1	21	190	119	1	331
	2	0	36	2326	263	2625
	3	0	0	126	1125	1251
	Total	63	238	2573	1389	4263

Measure of agreement – Kappa = 0.742

The Kappa figure for the evaluation schools (0.742) is very similar to that for the total trial sample (0.736) and the schools in the evaluation LEAs (0.742).

Table 2.15 shows that the proportion showing disagreement by more than one level is extremely low (less than 0.01%).

2) Reading

Table 2.16 shows the level of agreement for reading for the evaluation schools.

Table 2.16 Comparison of Teacher Assessment and test/task for schools in evaluation, 2003 data in reading

		Reading task/test				Total
		W	1	2	3	
Reading TA	W	89	15	0	0	104
	1	14	426	46	0	486
	2	0	52	2095	139	2286
	3	0	0	220	1162	1382
	Total	103	493	2361	1301	4258

Measure of agreement – Kappa = 0.806

The Kappa figures are 0.798 for the total trial sample, 0.797 for the evaluation LEAs and 0.806 for the evaluation schools, which are all 'substantial'/'near perfect' and very similar to each other. Again the proportion disagreeing by more than one level is very small.

3) Writing

Table 2.17 shows Kappa for writing for the evaluation schools.

Table 2.17 Comparison of Teacher Assessment and task and test for schools in the evaluation, 2003 data, in writing

		Writing task				Total
		W	1	2	3	
Writing TA	W	140	28	1	0	169
	1	46	377	78	0	501
	2	2	115	2542	152	2811
	3	0	0	99	683	782
	Total	188	520	2720	835	4263

Measure of agreement – Kappa = 0.768

For all children in the schools in the trial Kappa is 0.757 and for all in the evaluation LEAs Kappa is 0.767, both very close to the value of 0.768 for the evaluation sample. All are ‘substantial’ and all are similar. Again, very few show disagreement by more than one level.

It is also of note that the Kappa figures are very similar across the three subjects, although reading values are slightly higher, bordering on ‘near perfect’.

2004 data

In order to give a perspective on the robustness of the new arrangements, the Kappa values for 2004 are given in Tables 2.18-2.23. As schools in the trial reported to DfES (via their LEA) only their Teacher Assessments and not their test and task results, the 107 evaluation schools were asked to supply to the evaluation both Teacher Assessment data and task and test results. 76 schools sent a full set of data, and this provides the basis for the analyses.

1) Mathematics

Tables 2.18 and 2.19 give the levels awarded in 2004 in Mathematics.

Table 2.18 Comparison of Teacher Assessment and test/task for schools in evaluation, 2004 data, in mathematics

		Mathematics task/test						Total
		W	1	2C	2B	2A	3	
Mathematics TA	W	47	4	0	1	0	0	52
	1	12	153	17	2	1	0	185
	2C	0	18	430	46	4	1	499
	2B	0	0	43	590	75	4	712
	2A	0	0	4	29	655	65	753
	3	0	0	0	0	14	838	852
	Total	59	175	494	668	749	908	3053

Measure of agreement – Kappa = 0.856

Table 2.19 Comparison of Teacher Assessment and test/task and Cohen's Kappa for schools in evaluation, 2004 data, with level 2 collapsed, in mathematics

		Maths task/test				
		W	1	2	3	
Maths T	W	47	4	1	0	52
	1	12	153	20	0	185
	2	0	18	1876	70	1964
	3	0	0	14	838	852
	Total	59	175	1911	908	3053

Measure of agreement – Kappa = 0.911

The 2004 Kappa value of 0.856, with 0.911 for the 'collapsed' sample, compared to that for 2003 of 0.742 is a reassuringly strong increase.

2) Reading

Tables 2.20 and 2.21 give the levels for reading for 2004, on which Kappa was calculated.

Table 2.20 Comparison of Teacher Assessment and task/test for schools in evaluation, 2004 data, in reading

		Reading task/test*						Total
		W	1	2C	2B	2A	3	
Reading TA	W	60	0	1	0	0	0	61
	1	8	525	19	2	1	1	283
	2C	0	14	366	28	3	2	413
	2B	0	2	28	592	43	1	666
	2A	0	2	2	37	628	28	697
	3	0	1	1	0	72	827	901
	Total	68	271	417	659	747	859	3021

*In Reading, the data are given as a test and/or a task result. Again the data have been combined to give one level for each child in the data set, taking the higher level where the task or test level differs.

Measure of agreement – Kappa = 0.875

Table 2.21 Comparison of Teacher Assessment and task/test for schools in evaluation, 2004 data, with the level 2 collapsed, in reading

		Reading task/test				
		W	1	2	3	
Reading TA	W	60	0	1	0	61
	1	8	252	22	1	283
	2	0	18	1727	31	1776
	3	0	1	73	827	901
	Total	68	271	1823	859	3021

Measure of Agreement - Kappa = 0.907

Again the level of agreement has risen under the trial system.

3) Writing

Finally Tables 2.22 and 2.23 give the comparative levels for writing.

Table 2.22 Comparison of Teacher Assessment and task for schools in evaluation, 2004 data, in writing

		Writing task						Total
		W	1	2C	2B	2A	3	
Writing TA	W	121	4	0	0	0	0	125
	1	10	281	15	2	1	1	310
	2C	0	52	549	31	4	1	637
	2B	0	3	63	698	57	4	825
	2A	0	2	1	36	562	44	645
	3	0	0	1	1	20	493	515
	Total	131	342	629	768	644	543	3057

Measure of agreement – Kappa = 0.856

Table 2.23 Comparison of Teacher Assessment and task for schools in evaluation, 2004 data, with the level 2 collapsed, in writing

		Writing task				Total
		W	1	2	3	
Writing TA	W	121	4	0	0	125
	1	10	281	18	1	310
	2	0	57	2001	49	2107
	3	0	0	22	493	515
	Total	131	342	2041	543	3057

Measure of agreement - Kappa = 0.894

Again the level of agreement has risen under the trial system.

Tables 2.15-2.23 show that in all subjects there is a very high level of agreement between task and test result and Teacher Assessments, in both the current system and the trial, and where there is disagreement it is in most cases by only 1 level, and in many cases here, by just one sub-level. The Kappa data are summarized in Table 2.24, and emphasizes the increased level of agreement under the trial system, as expected, although the increased Kappa values may, to an extent, also reflect the decrease in level 1s and level 3s reported earlier.

Table 2.24 Summary of Cohen's Kappa measure of agreement for 2003 and 2004 for schools in evaluation

	2003	2004	2004 collapsed
Mathematics	0.742	0.856	0.911
Reading	0.806	0.875	0.907
Writing	0.768	0.856	0.894

This level of agreement between test and task results and Teacher Assessments was reflected in comments from questionnaires and from interviews. Both headteachers and teachers were asked about the extent to which there was variation between test/task and Teacher

Evaluation of the trial assessment arrangements for key stage 1

Assessment. The tables below show that the great majority said ‘few or none’ or ‘some’ and in very few instances was it ‘many’ or ‘all or almost all’.

Table 2.25 Headteacher perceptions of variation between Teacher Assessment and test/task

	Few or none	Some	Many	All or almost all	Don't know	Median/mode
Reading	42 (67)	15 (24)	2 (3)	2 (3)	1 (2)	Few or none
Writing	42 (67)	17 (27)	1 (2)	2 (3)	1(2)	Few or none
Maths	44 (70)	15 (24)	1 (2)	2 (3)	1 (2)	Few or none

Table 2.26 Teacher perceptions of variation between Teacher Assessment and test/task

	None or almost none	Some	Many	All or almost all	Median/mode
Reading	52 (59)	34 (39)	1 (1)	1(1)	None, or almost none
Writing	41 (47)	42 (48)	3 (3)	1 (1)	Some
Maths	50 (57)	32 (36)	5 (6)	1 (1)	None or almost none

Summary of variation between Teacher Assessment and test and task results

- The Kappa values for the agreement between Teacher Assessment and task /test levels were very high in 2003, in all subjects.
- In 2004 the values were even higher, which is to be expected since in 2004 in the trial teachers were required to take the test / task result into account when arriving at a Teacher Assessment, and so the two measures are less independent.
- That the level of agreement has not been reduced strongly suggests that the robustness has not been adversely affected.

2.2 Training and moderation

Another way in which the robustness of the assessment arrangements in the trial can be evaluated is in the training and moderation that has been provided for schools in the LEAs of the evaluation sample, which informed the process of reaching Teacher Assessment judgements in those schools. The new arrangements give a great deal of discretion to individuals who are physically very dispersed. The necessary consistency and rigour in judgement relies on high quality training and effective moderation that in some way reaches all the participants.

This section will consider:

2.2.1 Training

2.2.2 Moderation

2.2.1 Training

LEA training observations

Three LEA training sessions were observed; one in each of three separate evaluation LEAs. The number of teachers present varied from about 25 to double this number. The most obvious commonalities between the training events were clear statements of the substantive changes in the requirements for end of key stage 1 reporting. These were typically plenary presentations of 'the facts' supported by paper-based versions of information for individual teachers. This took about three-quarters of an hour to an hour in each case.

Unsurprisingly, variations appear in the guidance and advice offered to teachers in how to respond to these changes. This is not only a reflection of the differing approaches of the individuals offering the guidance, but also of the alertness of teachers in envisaging problems that might arise, in an as yet untried arrangement and the confidence or willingness of teachers to raise points in a meeting. Each session observed was one of a repeated series by which means LEAs expect to train all affected teachers. It is most likely that subsequent 'repeat' sessions will have been modified by the experiences of the sessions that were observed.

The effects of the changes were broadly characterised in terms of:

- reducing teachers' workload;
- simplification of reporting results;
- offering much greater flexibility for teachers in how they manage assessment at KS1;
- offering opportunities for teachers to use and develop their professional skills.

On all occasions reference was made to documentation emanating from QCA such as exemplar materials of pupils' work characterising National Curriculum levels and Assessment Focus literature.

The least satisfactory approach, perhaps, was to offer advice on a take it or leave it basis: 'you can do this or you can do that or you can do neither' without leading or at least offering a discussion of what might be the most productive and professionally worthwhile course of action for the teacher to follow. This could be characterised as the freedom without too much responsibility approach. In one venue there was an offering of rather generalised teaching ideas for the classroom – all potentially useful in themselves, but not linked very clearly to how they might be used in this particular assessment context. Other trainers were much more detailed in offering sources of useful material and suggesting modes of use.

There was a common theme of encouraging teachers to take advantage of the opportunity to develop their professionalism and not take the 'easy way out' of simply running tests or tasks and using the results from these as Teacher Assessment. One of the pitfalls of encouraging authentic Teacher Assessment is in how it is to be managed without it becoming more of a burden than the existing arrangements. Clear advice was given in all training sessions that extensive portfolios of pupils' work were not required and that whatever assessment records were kept were at the teacher's discretion for their own needs (though test / task results must be recorded). Such material might well be useful for moderation purposes, but it was not a requirement.

In one venue teachers revealed their anxiety concerning the variation that can occur between test results, task results and Teacher Assessment of pupils. Teachers were encouraged to see the proposed changes in terms of seeking better quality assessment and that whilst their

Evaluation of the trial assessment arrangements for key stage 1

Teacher Assessment should take account of the test result (for example), there was nothing wrong in offering a Teacher Assessment that differed from it, provided that teachers were confident in their own procedures.

Overall there was a positive atmosphere at the training sessions, with the sentiment that this is a potentially better arrangement than hitherto, so we must work to ensure its success.

Findings from questionnaires and interviews

The great majority of teachers seem to have attended some or all of the training provided by the LEA.

Table 2.27 Teachers' responses to 'Have you attended (or will you be attending) your LEA's assessment/moderation training sessions for Year 2 teachers?'

	All	Some	None	Don't know
Attend training	62 (54)	39 (34)	11 (10)	3 (3)

Heads, teachers and LEA coordinators were asked via questionnaire about their views on previous training, training for the trial, its usefulness and quality. Heads and teachers both generally felt that in 2003 they had been 'well' supported in terms of information, training and guidance received from their LEA (heads) and their school (teachers).

Table 2.28 Responses from teachers and headteachers to 'How well-supported was the school last year in terms of information, training and guidance received from:

	Very poorly	Poorly	Adequately	Well	Very well
the LEA?' (Heads) (n=79)	0 (0)	0 (0)	25 (30)	37 (45)	17 (20)
the LEA?' (Teachers) (n=91)	0 (0)	0 (0)	31 (27)	39 (34)	35 (30)
your school?' (Teachers) (n=92)	3 (3)	1 (1)	40 (35)	42 (37)	20 (17)

Heads and teachers were also asked (in their second questionnaire) about the same issues under the trial system in 2004.

Table 2.29 Responses from headteachers and teachers to 'How well-supported was the school this year in terms of information, training and guidance received from:

	Very poorly	Poorly	Adequately	Well	Very well
the LEA?' (Heads) (n=63)	1 (2)	4 (6)	30 (48)	18 (29)	10 (16)
the LEA?' (Teachers) (n=87)	0 (0)	1 (1)	36 (41)	25 (28)	24 (27)
your school?' (Teachers) (n=87)	0 (0)	2 (2)	21 (24)	37 (42)	27 (31)

The information, training and guidance received from the LEA was not received quite so well and is signified by a median/mode of 'well' in 2003 to 'adequately' or 'well' in 2004. When asked for reasons to support their views one head suggested "*training has been rushed and inadequate*".

Evaluation of the trial assessment arrangements for key stage 1

A teacher also expressed concern about the timing:

Information, training and materials needed to be in schools in September NOT at the end of MARCH.

(teacher)

LEA coordinators acknowledged that some of the training may have taken place too late and that this in turn may have meant that some teachers did not use the tests/tasks as flexibly as they might have.

In the interviews (and it arose also in a couple of the focus groups) teachers and LEA coordinators were asked about their experience of training. Seven teachers were positive about the training, and two were very positive. Two were fairly neutral and just one found the experience a negative one. Those having a positive view about the training said that it helped by making it clear what needed to be done, to hear what other schools were doing, and it was helpful to receive the exemplification materials. The two neutral teachers appear to have attended the same training session and felt it emphasised the past a little too much, and the negative teacher felt that not enough guidance was given. Teachers generally felt that the time given to attending training was worth it. A teacher who could not attend training due to ill-health said that it was difficult to get her 'head round' Teacher Assessment. Again teachers said it would have been helpful if the training had taken place earlier.

Seven of the LEA coordinators mentioned training and were asked how they felt the training had been received and how successful they thought it had been. Of these seven, five were positive claiming that some good ideas were passed on, and some misunderstandings cleared up, although it did entail a lot of work. The other two were neutral, one claiming that not all teachers understood the key messages, and another saying that it could have been better both by timing and by more specific advice.

Overview of training

All schools need training on the assessment arrangements, and the positive feedback about training in the trial suggests that what was done was helpful to at least some degree and should, under 'roll-out', be available to *all* Year 2 teachers and managers. In order to maximise the help, it seems advisable for different training to be offered to 'lone' Year 2 teachers than to those working in larger schools. It might also be helpful if different training were to be offered to teachers new to Year 2 compared to those with experience of the existing arrangements (since comparisons between what is new and what went before is helpful to one group but distracting to the other). It is also likely that different training is needed for school managers compared to Year 2 teachers, with the former focused mostly on the overall system and possibilities for using the available flexibility and for in-school moderation, but with the latter focused more directly on levelling and quantity of evidence. It also seems important to have some training, especially for management, early in the year so that the school can begin to set up appropriate systems as soon as possible.

Overall the most desirable quality in the training was felt to be clarity – clarity about the need for evidence; about what the flexibility really means (the choices you can make, those you cannot, and the implications of each of the choices); clarity about levels and grading within them; and how to decide on a level in each subject (which does vary between them).

In the trial there was some evidence that in the earliest training the simple message of flexibility was sometimes overplayed at the expense of clarity about what would actually be involved. In 'roll-out', given that those schools and LEAs who have not been part of the trial will be 'starting from the beginning' and inevitably will begin with a similar uncertainty

Evaluation of the trial assessment arrangements for key stage 1

about what will be involved, it seems important not to over-emphasise the flexibility aspect, but to try to give clearer guidance from the start. All of those who have been involved in the trial have become clearer about the nuts and bolts of the initiative as a result of the experience, and so may be helpful sources of clarity.

Summary of training

- The majority of teachers attended some or all of the training provided by LEAs.
- Most teachers felt they benefitted from the training they received.
- Training should be made available to all teachers, with those teachers new to Year 2, and / or in single entry schools (with no other Year 2 teacher) receiving additional training where necessary.
- Some training should take place early on in the school year.

2.2.2 Moderation

Moderation of assessment results is seen by headteachers and LEA personnel as the main mechanism to ensure rigour and consistency. The following summarises the key findings related to moderation, using evidence from focus groups, interviews and questionnaires (to headteachers, teachers, LEAs) conducted during the evaluation.

The plans for moderation submitted by trial LEAs

Plans for moderation were submitted by all LEAs in the trial. Appendix E lists summaries of the plans received, separated into those from evaluation LEAs and those from the other trial LEAs.

The vast majority of LEAs planned to moderate around a quarter of their schools, with schools arranged on a rolling programme such that they are subject to moderation every four years. In addition to those schools due to be moderated, additional schools causing concern were also to be moderated. This was with the exception of one LEA which intended to moderate all schools.

The number of key stage 1 schools within evaluation LEAs varied between 39 and 437 (and between 29 and 300 for trial schools not in evaluation). The number of schools to be moderated varied between 10 and 115 for evaluation LEAs, and between 8 and 120 for non-evaluation LEAs.

The number of moderators involved in the process varied according to the number of schools to be moderated. Most LEAs utilised between 4 and 20 moderators, usually a mix of LEA personnel (advisors, primary strategy consultants, school improvement officers) and seconded teachers (Year 2 teachers, deputy headteachers, assessment coordinators, retired headteachers).

All LEAs planned training for moderators and this typically ranged from 2 half day sessions to 2 full days. The planned moderation activities varied between LEAs, with most moderators visiting schools in June, but some visits were as early as May, and one LEA planned to visit in March.

The moderation visits were planned to include discussions with Year 2 teachers, assessment coordinators and/or headteachers and were to be centred on the moderation process, how the school organised the assessment and its requirements and any difficulties experienced. Central to the intended moderation visit was the observation of children and the discussion of variations between Teacher Assessments and task/test results. In addition, scrutiny of work

Evaluation of the trial assessment arrangements for key stage 1

was planned, to include pupils' books, displays, pupils' records, and task and test scripts/records. Amongst LEAs, the minimum sample size for moderation in each school was fairly standard at six children. These were planned to be children of varying ability, as well as those where the task/test and Teacher Assessment levels differed. For one LEA, the plan was for the moderator to pre-select children using results for the Year 2 classes forwarded by the teacher.

Feedback was predominantly planned to be given on the day, both verbally and also through a brief moderation report. This was to be followed with an evaluation of the moderation as a whole as well as any feedback given to the LEA on the process, and was to be culminated in a report sent to the school/s involved. One LEA planned to provide a certificate of validation for the key stage 1 results to each school at the end of the moderation visit.

In addition to visits to individual schools, LEAs planned to organise moderation meetings for clusters of schools as an opportunity for Year 2 teachers to bring assessment samples for moderation judgements. These were planned for the Spring and Summer terms. Some LEAs planned a postal moderation of writing scripts as well as a 'surgery' for teachers to bring marking and levelling queries to an LEA consultant on an informal basis.

One LEA was unusual in that it split the 25% of schools to be moderated into two. Half the schools were to receive a one-to-one moderation visit which was planned for the moderation activities stated above, as well as to watch test/tasks. The other half of schools formed small groups and met with a moderator to discuss evidence from Teacher Assessments and task/tests.

LEAs planned to consider any disagreement between schools and moderators concerning pupils' results in similar ways. The intention was generally that any differences of opinion would hopefully be resolved during the school visit, but for those not resolved, initially the lead moderator would make an additional judgement, and if agreement still was not reached, an independent moderator from a neighbouring LEA may be utilised.

General impressions of the implementation of the moderation

It is clear from the evidence from interviews, questionnaires and focus groups that teachers responded well to the moderation process, which they found to be positive and supportive. The very large majority of teachers who commented on moderation reported a positive experience. Some comments were:

moderation was great

(teacher)

they [the moderators] were with us

(teacher focus group)

NQTs have found it really helpful [for] supporting them in their judgement"

(teacher focus group).

However, one teacher reported having had a negative experience, saying that she had felt bullied into changing her 2As in maths to 3s and felt that the whole moderation experience was like OFSTED coming in.

Some schools needed reassurance that the moderation process was designed to be supportive and there were anxieties that the process would be inspectorial. LEAs commented that moderation is about "*ensuring quality of teacher judgements – not about checking the levels are right.*" This raises the issue of the tone of moderation. It was felt quite strongly that it

Evaluation of the trial assessment arrangements for key stage 1

should be a process where teachers are playing an equal part and not having decisions forced upon them.

The comments made by LEA assessment coordinators in interviews were also very positive. They felt that whilst there had been some difficulties, these were to be expected in the first year of a new system, and that overall the process went very well. The LEA coordinators interviewed all felt that in some cases teachers were confused and anxious to begin with, but contact with the LEA helped to reassure and inform teachers, and moderation itself was useful to get a good feel for the school and to inform future work.

LEA coordinators highlighted teachers overcoming their reliance on tests/tasks as one of the main issues of moderation, and attributed it to lack of teacher confidence in their own judgement. In relation to this, several teachers reported that moderation contributed to their own confidence in their ability to do Teacher Assessment, partly because the moderators had agreed with what they had done.

However, the non-uniformity of the moderation process from LEA to LEA was commented upon by some coordinators. Whilst this allowed choice and flexibility, which was valuable to design a process to suit local circumstances, further guidance as to what constituted good practice would be helpful in ensuring some consistency nationally.

Training moderators

In many LEAs, moderators included auditors from previous years and thus, had experience of visiting schools to audit tasks/test administration and outcomes. Inevitably however, there were moderators completely new to the role and indeed some LEAs took the opportunity of the trial to revamp their systems and employ new moderators. Some LEAs commented positively on employing Year 2 teachers as moderators as *“they not only have a lot to offer to the process, but they also take a lot from the experience which can benefit their school”*. Regardless of the makeup of the team of moderators, training was seen to be vital to the team’s competency and the effective implementation of the moderation process.

Guidelines regarding a moderator’s role were also seen as paramount and it was felt that appropriate training should encompass these. Likewise clear information on the moderation process and how to deal with certain situations was part of effective moderator training. Some LEAs conducted an exercise at the start of the moderation period whereby the team of moderators came to a shared sense of levels, using material in a postal audit. This was said to have helped in ensuring consistency in the ensuing process.

The quantity of evidence for moderation

LEAs felt that it was important to reinforce to teachers that the moderation material should be material already collated and not additional material especially done for moderation. LEAs felt that teachers should not utilise portfolios, instead, during moderation, pupils’ work and records should be examined. On-going records that reflect levels make the moderation and the justification easier, but require more work (or a change in focus in the work).

However, LEAs reported that teachers were unclear and anxious about the evidence required for moderation. There was a wide variation in the amount and type of records kept. Some teachers were over-conscientious with the amount of evidence whilst others provided insufficient firm evidence or too much inappropriate evidence during moderation visits. It is evident that some teachers felt that more needed to be collected for moderation than for their usual assessments, probably due to the need to convince someone else of the judgment. Whilst it is important that teachers are informed about the amount of evidence to supply, it is even more critical that they have a good understanding of the type of evidence that is appropriate. Consequently, training for teachers needs to provide clear models of the evidence that teachers could provide for moderation.

Table 2.30 The amount of evidence collected for moderation purposes

	Didn't collect any	One piece	Several pieces of work	Other*
Writing	6 (7)	5 (6)	61 (69)	16 (18)
Mathematics	10 (11)	3 (3)	54 (61)	21 (24)

*The most common response under 'other' was the use of pupils' exercise books.

As Table 2.30 shows, schools to be moderated had more teachers that collected 'several pieces of work'.

An important element in deciding whether a piece of work that is 'at' a particular level, is 'worth' that level is how much help the child received in doing it. Moderators are less well placed than teachers to judge the amount of help that a child received in completing a piece of work (as they were not there), but teachers can struggle to identify the help that they actually give, as much of it is incidental rather than planned. Nevertheless, moderation visits are an opportunity to be able to offer additional information and insight that might clarify or resolve some differences in judgements that remain after the examination of work.

The process of moderation

An emphasis within the moderation process on teacher judgement having precedence over test result, enhances teachers' sense of professionalism, but also increases their workload, as they feel the need to justify their judgements. This pressure comes mostly from an awareness that test results are trusted more than teachers are.

...also in the past when Teacher Assessment has been reported alongside a test result I have a feeling that parents see the test result as the accurate one... [other teachers concur].. and if the test result says 3 and the Teacher Assessment says 2A in the past that's only proved that the teacher was wrong because in parent's eyes... I would suggest in parent's eyes the test must have been the accurate one... and that is certainly true at key stage 2 now where no matter what the teacher says actually it's the test results isn't it... because that's the real one but ... that's not the case... it's the teacher assessment which is the most accurate and you said 2A right up to the day of the test... the test said 3 but you're still a 2A... and is tomorrow morning a 2A...

(teacher)

Moderation which focuses on talking to teachers about their evidence and reasons can contribute to that pressure, and reinforce the teachers' impressions about trust. This is a complex situation. Do levels apply to pupils or their work? Teachers know pupils, whereas tests can only know their work. Pupils' work and test performance can be interpreted in many ways.

I've had a few children in my class whereby sometimes they will be silly throughout the year... yes... and the type that obviously are bright... but they just don't want to bother... and they've left it until these tests and obviously this one child he was a 1 because we've done previous tests with him all throughout the year... a 1 throughout...and then he's really put his mind to it and he's come out with a clear 2... and obviously my Teacher Assessment would be a 1... whereas in his test he's done a lot better...

...well obviously I've talked to my colleagues and I mean we're all in agreement that this child just hasn't bothered but he doesn't show it in his work and he's left it until

Evaluation of the trial assessment arrangements for key stage 1

then to show us... but overall if you're thinking about for the whole year from September until now no...

(teacher)

...you don't normally be performing at 2B in my opinion and then suddenly get a 3 in a test because then that makes me... I'm obviously underestimating that child... it's the other way normally that they perform really well or okay in class and then on that piece of work... obviously off ...or it doesn't suit them. .. the writing, yes ... I don't want to write about a princess ... I don't want to do it.... I want to write about football... everybody does it.. it doesn't mean to say you're a bad writer...

(teacher)

During some moderation visits, the need for funding to support moderation work within schools was highlighted. This, it was felt, could secure non-contact time for teachers to do the work, rather than relying on teachers' own time and goodwill.

Training meetings were felt to provide a good opportunity for teachers to network with other teachers, which can often be an excellent source of information and support during moderation.

Moderating in School

Several LEA coordinators felt that small schools with lone Year 2 teachers would present greater challenges because of the lack of an opportunity for different Year 2 teachers in a school to work together, cross moderate and discuss approaches. They felt that teachers new to Year 2 in single form entry would be at an especially high risk of being overwhelmed, where in other schools the established Year 2 teacher would be able to support the new. This was linked to whole school and cross-school moderation, and although head teachers did not express any concerns about single Year 2 teachers, many referred to the need for whole-school and cross-school moderation.

Where there is only one Year 2 class within a school, the teachers are said to be less secure in their judgements. Whilst training can play a large role in this, one suggestion that arose on many occasions, was to moderate across the school (with Year 1 and Year 3 teachers). Within some schools different teachers moderated each other's judgements and held meetings (in pairs) to discuss levelling and this was reflected in the accuracy of judgements. Involving other teachers at school level spreads the responsibility of the work and is good CPD for teachers.

... and one of the things that really impressed me was that in one school I went to I was asking them... this teacher about the moderation of the writing... and she said that she'd taken on board the comments I'd made at the training about liaising with other colleagues in another small school because they don't see the range... and she said actually she said about five of us got together five different schools had got together and organised their own cross moderation meeting... and I think that is a really positive thing... if people can see the benefit of that... supporting each other and seeing standards in that way I think that was a really good thing that came out of it... and I don't think they would have done that if it hadn't been for the trial... I think it's given them that feeling of autonomy which is very positive...

(LEA coordinator)

This practice does, however, have implications upon the workload of teachers and may indeed only be practicable in schools with two Year 2 classes. The teacher who had a negative experience of moderation is a lone Year 2 teacher and it has led that LEA to put in place termly meetings where all lone Year 2 teachers can meet with a moderator (a lone Year 2

Evaluation of the trial assessment arrangements for key stage 1

teacher) to discuss issues around assessment and moderation. Another suggestion is to have cluster moderation meetings as part of the process. Another option, carried out in some LEAs is the adoption of ‘surgeries’ for teachers who want support with levelling; these however “*tend not to be well attended*”.

Other issues

LEA coordinators found that a whole school approach to assessment had a significant effect upon the teachers within a school. Where there was a clear Assessment Policy that all staff knew and understood, teachers had confidence in the judgements made. In schools where the Year 2 teacher appeared to be finding his or her own way forward, teachers’ confidence and indeed moderators’ confidence in the judgements made was lessened. Levelling in maths was highlighted by several LEAs as one of the main issues in moderation of teachers’ assessments. This was due to the introduction of sub-levels in mathematics this year. It was specified as an area for further training and additional support.

One trial LEA chose the children for moderation, rather than leaving it up to the teacher. They felt this eased the burden as teachers couldn’t collate extra work for moderated children as they didn’t know which children were going to be moderated. The children chosen were those who had discrepancies between Teacher Assessments and tests. This was felt to have some drawbacks however, and perhaps LEAs should be interested in moderating children where the teacher was not so sure about differences attained in tests and Teacher Assessments.

The pressure of time in moderation leads to a focus on where the test result differs from the final assessment. This could allow many cases to go unnoticed where the final result is at odds with the class work (but agrees with the test result), so that moderation would not be a check on the distortion of the assessment by test preparation. It also implicitly gives tests more status than teacher assessment based on classwork.

Views about moderation

The value and contribution of moderation was widely appreciated by teachers, with just a few exceptions that seemed to reflect differences in interpretations of levels (often arising it seemed from the tension between ‘best-fit’ and coverage). This was despite the common feeling that the workload for moderated schools was greater than that for others because of the collection and organisation of evidence for external moderation purposes that exceeded or conflicted with the approach used within the school.

...it’s taught us how to be more accurate and more objective with teacher assessment... as the reliance was more on teacher assessment it made us scrutinise children’s work more and not just look at it overall, but unpick it particularly in writing...

(teacher)

... it was useful having the people from the LEA come because it was I remember there was one case where a writing... a piece of writing one of my children had done... I’d given her a 2A... but the spellings were not decipherable and then we had a discussion he was saying to me that here was a person who didn’t know her... if he couldn’t understand what she had written then I had to understand that then wasn’t a 2A... because what happened knowing the children and knowing how they spell sometimes not well... oh they can do this it’s just the spellings... we use another ...so it was useful having the LEA people come and really standardise it... it made it quite clear for us that it’s 2A potentially... but because they are still not spelling correctly we can’t give them that grade... so that was useful having that... what else have we – how else do we do it... it was nice looking at it from somebody who doesn’t know them...

(teacher)

Overview

The trial has not provided clear evidence of identifiable ‘good practices’ and ‘poor practices’ in terms of moderation systems. Different LEAs adopted different approaches in the trial, and in every case they were, at the end of the year, identifying refinements and improvements for the following year, but without dropping any elements entirely as a perceived mistake. The varied circumstances of each LEA, including differences in size, accessibility, school patterns, socio-economic factors and so on mean that each LEA requires some flexibility to shape a system to suit them, which can be evolved over time in the light of experience.

That said, elements that were either very common or were viewed particularly positively were:

- early meetings of moderators to ensure consistency within the LEA (which might be considered also in cross-LEA contexts);
- an emphasis on negotiation rather than imposition;
- opportunities for moderators to meet the children whose work is being moderated, and not just deal with pieces of work in isolation (although this must be handled with care);
- training that is clear about the evidence that is needed for moderation.

A further characteristic of success in LEA moderation seemed to be:

- the availability of LEA personnel to follow up ‘loose ends’ of various kinds, including offering additional support to those that need it (for example teachers new to Year 2 who are not in any kind of support network within their own school).

Wherever possible, encouragement to schools to develop internal moderation procedures, through which teachers come to shared understandings about levels of work and working, and discuss cases and evidence to assist in the assessment judgements, is better than LEAs needing to offer additional support to individuals. Nevertheless LEAs still need to understand their schools and teachers well, in order to know where targeted support remains necessary.

Summary of moderation

- Successful completion of moderation can help support the confidence of teachers in their ability to carry out the Teacher Assessments.
- Most moderation experiences were positive and teachers felt supported.
- LEA coordinators felt the moderation sessions were successful.
- At its best, moderation should be supportive of teachers and not have the atmosphere of an OFSTED inspection. Judgements can be challenged in positive and supportive ways.
- Teachers were unsure about how much evidence to collect for moderation, and what would be inappropriate. Firm guidelines are needed, although less confident teachers may feel the need to collect more.
- ‘Lone’ teachers in small schools may need extra support for moderation.

2.3 Other evidence to evaluate the robustness of the data from the trial system

2.3.1 A comparison of evaluation school data and DfES data.

2.3.2 The confidence of headteachers and LEA assessment coordinators.

2.3.1 Evaluation school data and DfES data - a comparison

Robustness depends in part on the quality of information transfer within the system. The data sent to LEAs and on to schools about pupil levels should accurately reflect the judgements of the schools, and be the same as that kept in school records and sent home to parents.

The evaluation did not have direct access to these data in order to make comparisons. However, Teacher Assessment levels on the children in 76 schools in the evaluation sample were obtained by the evaluation team from two different sources. Firstly the levels returned by the schools to the LEA and DfES were forwarded to the team. Secondly, the schools submitted the levels directly (prompted by the need to have test and task data as well, which were not returned to LEAs or DfES). Each data set was checked for equivalence in broad terms – the number of children at each level – and differences were found.

Speaking and Listening

There was disagreement in 26 of the 76 schools. In 21 of these schools the disagreement involved just one child. In one school it involved 2 children, and in one school it involved 4 children.

Reading

There was disagreement in 35 of the 76 schools. In 22 of these schools the disagreement involved just one child. In five schools it involved 2 children, in four schools it involved 3 or 4 children, and in one school the results looked quite different.

Writing

There was disagreement in 34 of the 76 schools. In 22 of these schools the disagreement involved just one child. In four schools it involved 2 children, in four schools it involved 3 to 5 children, and in two school the results looked quite different.

Mathematics

There was disagreement in 35 of the 76 schools. In 19 of these schools the disagreement involved just one child. In four schools it involved 2 children, in four school it involved 3 to 5 children, and in five schools the results looked quite different.

The inaccuracies are unlikely to have had a significant effect on the analyses of the evaluation, since in over half of the cases where the disagreement involves just one child it is due to one of the data sets having an additional child, and this does not affect the variation analysis at all, and where there is, for example, one fewer child at level 3 and one more at level 2A, which may have affected the variation analysis, the number of children involved is very small.

Nevertheless the disparities are of concern. There are a number of possible causes of the disagreements in the sources of data received by the evaluation:

- the possibility that LEAs are receiving different data from that which the parents receive;

Evaluation of the trial assessment arrangements for key stage 1

- the possibility that where levels have been changed as a result of moderation, this is reported to the LEA but not to the parent;
- the possibility that the person who gives the Teacher Assessment, i.e. the teacher, is not the person who compiles the data for the LEA and that the data compiler does not recognize errors always, due to unfamiliarity with the children;
- the possibility that there are errors at data input level that remain uncorrected.

2.3.2 The confidence of headteachers and LEA assessment coordinators

One final measure of the degree of accuracy in assessment actually achieved in the trial, and therefore of the robustness of the system, is the confidence that headteachers and LEA personnel have in the results. This was addressed in the LEA questionnaire item: How confident were you about the accuracy of the overall attainment levels produced by teachers? And in the Headteacher questionnaire item: How confident are you about the accuracy of the overall attainment levels produced? (separated into reading, writing and mathematics).

Tables 2.31 and 2.32 show the responses to these questions.

Table 2.31 Confidence of LEA coordinators in overall judgement levels (Teacher Assessments) in 2004

	Very confident	Quite confident	Not very confident	Not at all confident	Don't know	Median / mode
LEA	8 (36)	14 (64)	0 (0)	0 (0)	0 (0)	Quite confident

Table 2.32 Confidence of headteachers in overall judgement levels (Teacher Assessments) in 2004

Headteacher	Very confident	Quite confident	Not very confident	Not at all confident	Don't know	Median/mode
Reading	41 (65)	22 (35)	0 (0)	0 (0)	0 (0)	Very confident
Writing	37 (59)	24 (38)	2 (3)	0 (0)	0 (0)	Very confident
Maths	39 (62)	21 (33)	3 (5)	0 (0)	0 (0)	Very confident

These results suggest that the overall attainment levels are felt to be accurate in the large majority of cases.

The views expressed in interviews and focus groups also tended to show that the great majority of stakeholders felt that teachers' judgements would be accurate. This included all of the LEA coordinators interviewed. However, some teachers and some members of parent focus groups had a degree of uncertainty about this.

2.4 Key issues and findings

- In comparing 2003 with 2004 data, the Teacher Assessment means are higher in 2004 than in 2003, whereas the test/task means are lower in 2004 than in 2003.
- In 2004, there was an increase in the proportion of level 2 in Teacher Assessments, and a decrease in level 1 and level 3.
- All the differences (even where statistically significant) are small and many of the effects could be subsumed within the degree of change that one may expect from year to year.
- The differences found from 2003 to 2004 were also found in speaking and listening, which was not directly affected by the change to assessment arrangements.
- The differences may represent an increase in the accuracy of reported assessment.
- Neither the gender of child nor the LEA in which the school is located appears to have a significant effect on the trial outcomes. However, age does appear to have some effect on Teacher Assessment results.
- Concerns about Infant/First/Primary schools having differences in assessment outcomes as a direct result of the trial seem to be unfounded.
- The extent of agreement between test/task and Teacher Assessment in 2003 were very high. The level of agreement was higher in 2004, which is to be expected since the two measures are less independent than in the existing system. This continued high level of agreement with no reduction suggests that robustness has not been reduced.
- Most teachers attended some or all of the training offered and the majority found it was helpful.
- Training should be made available early and to all teachers.
- Differentiated training should be made available to some groups of teachers, especially lone Year 2 teachers and teachers new to Year 2.
- Most moderation experiences were positive; teachers felt supported and LEAs felt the sessions were successful.
- Moderation should involve negotiation rather than imposition. Judgements can be challenged in positive and supportive ways.
- Experiencing and completing the moderation process appears to bring an added confidence to the teachers involved.
- LEA coordinators and headteachers have confidence in the accuracy of Teacher Assessments.
- Our findings suggest that the outcomes of the trial assessment system are likely to be as robust as the outcomes of the present system.

3. Workload and management implications of the new arrangements

All change comes at a price. New procedures which promise simplification, less work and so forth, still have to be mastered and embedded into new routines; this implies learning and requires time and effort. A significant concern at the outset about the new arrangements for assessment at key stage 1 was that however well meant or desirable they may be, they would inevitably add to the workload of teachers and schools at least in the short term and have implications for the management of the assessment process. This section looks at the pre-trial expectations and the actual trial outcomes in relation to workload and management.

The chapter is divided into three main sections:

3.1 Workload implications of the new arrangements for teachers, schools and LEAs

This section considers responses by Year 2 teachers, headteachers and LEA coordinators to questions about their own workloads and in the case of the latter two groups, their perceptions of the workload of Year 2 teachers. These responses have come from questionnaires, interviews and focus group discussions. There are three sub-sections:

Pre-trial (2003) experiences and perceptions of workload for teachers and schools in relation to KS1 assessment.

The impact of the trial (2004) on workload for teachers and schools

The impact of the trial on LEAs

3.2 Management implications of the new arrangements for schools and LEAs

This section describes the management implications of the new arrangements other than those directly concerned with workload. There are four sub-sections:

Issues preceding the trial

Issues following the trial

Overall impact on the new arrangements on school management

Impact of the new arrangements on management of assessment within LEAs

3.3 Key findings and issues

3.1 Workload implications of the new arrangements for teachers, schools and LEAs

The information is set out in three sub-sections:

3.1.1 Pre-trial (2003) experiences and perceptions of workload for teachers and schools in relation to KS1 assessment

3.1.2 The impact of the trial (2004) on workload for teachers and schools

3.1.3 The impact of the trial on workload in LEAs

3.1.1 Pre-trial (2003) experiences and perceptions of workload for teachers and schools in relation to KS1 assessment

Headteachers' and teachers' views on existing workload

Nearly all headteachers, and teachers, responded that the amount of work last year was 'reasonable', 'a large amount' or 'too much' with a substantial majority of both groups identifying 'large amount' (Table 3.1).

Table 3.1 - Headteacher and teacher questionnaire (1) – Amount of work in doing assessments

1. How much work was required in doing the assessments in 2003?	Minimal	Small amount	Reasonable amount	Large amount	Too much	No answer
Headteachers (n = 83)	0 (0)	2 (2)	18 (22)	50 (60)	10 (12)	3 (4)
Teachers (n = 115)	0 (0)	0 (0)	13 (11)	66 (57)	28 (24)	8 (7)

The two main reasons offered for the relatively heavy workload in 2003 and before were that:

- National Curriculum tests were too time consuming especially in the assessment of writing;
- Tests take up too much class time.

The way that National Curriculum tests use up a large amount of classroom time is exemplified in the following comment:

Too many hours of actual tests for the children. Too long non-teaching/testing time - time wasted for learning. New writing marking system made time long.

(teacher)

The writing component seems to have been one of the main reasons given for the heavy workload, as shown in the last comment and in others such as

The changes to the writing marking criteria in 2003 significantly increased the workload...

(teacher)

Even a generally positive comment noted the problems with writing:

I was very happy with 2003 SATs. I considered that tests and tasks were excellent assessments that provided me with an overview of my pupil's [sic] learning and enabled me to identify areas of learning which needed more attention next year. Writing task too onerous and complicated.

(teacher)

Expected impact on Year 2 teachers' workload from heads, teachers, LEA and governors

- Most teachers, heads and LEA coordinators did not think the workload would increase, with LEA coordinators being the most optimistic.
- Issues mentioned by heads/teachers that might increase workload were moderation, training and simply coping with a new system.
- Governors, where they commented on workload, were almost equally divided over whether it would increase or decrease, with a slight majority (by less than 5%) expecting an increase, although some thought it would increase only in the short term.
- The main reasons given for thinking it might increase were issues relating to training and paperwork. The main reason given for thinking workload might decrease was that there are fewer tests.
- The majority of teachers said they would be keeping several pieces of work in each subject. There was some concern from LEA coordinators that too much, insufficient or inappropriate evidence would be collected.

3.1.2 The impact of the trial (2004) on workload for teachers and schools

Actual impact on Year 2 teachers' workload

Over half the teachers reported the workload to be 'a little less' or 'about the same'. There is, perhaps not surprisingly, a shift to greater workload for those moderated in comparison to those not moderated. Approximately the same overall totals for 'a little less', 'about the same' and 'a little more' show a net migration from 'a little less' towards 'a little more' (Table 3.2).

Table 3.2 - Teacher questionnaire (2) – Comparing teacher workload

3. Compare the workload you experienced this year against that last year.	Lot less	Little less	About the same	Little more	Lot more	Don't know	No answer
Overall (n = 82)	5 (6)	22 (27)	23 (28)	16 (20)	10 (12)	6 (7)	6 (7)
Not Moderated (n = 27)	2 (7)	9 (33)	7 (26)	4 (15)	1 (4)	2 (7)	2 (7)
Moderated (n = 55)	3 (5)	11 (20)	15 (27)	11 (20)	9 (16)	3 (5)	3 (5)

Headteachers' views on Year 2 teachers' workload

Headteacher questionnaire responses broadly concurred with teachers' views, though with less deviation from the most typical opinion of 'About the same' (Table 3.3).

Table 3.3 – Headteacher questionnaire (2) – Views on workload of Year 2 teachers

4. Compare the workload of the Y2 teachers this year with that of last year (n = 63)	Lot less	Little less	About the same	Little more	Lot more	No answer
	1 (2)	11 (17)	34 (54)	10 (16)	6 (10)	1 (2)

LEA coordinators' views on Year 2 teachers' workload

LEA personnel saw the situation of Year 2 teachers in a better light than the teachers themselves, with the largest number considering there to have been a small reduction (Table 3.4).

Table 3.4 - LEA coordinator questionnaire (2) – Impact on teacher workload

3. What impact did the assessment trial have on the workload of Y2 teachers (n = 22)?	Large reduction	Small reduction	No change	Small increase	Large increase	Don't know	No answer
	0 (0)	9 (41)	5 (23)	4 (18)	0 (0)	1 (5)	3 (14)

There were several comments that there would be an inevitable small increase in workload but that this would be acceptable, as it would, if things were done properly, facilitate a more enlightened approach to assessment:

Teachers felt OK about a small increase in workload this year as they see it as a short term pressure. They are confident that next year, if the system continues workload would be less because they will be more aware of what is required from September. They will be more focussed on formative marking as an ongoing process. Teachers have said that the assessment process was less pressurised with less testing to do.

(LEA coordinator)

Interview and focus group results

By their nature, individual interviews and group discussions reveal greater complexity of views and wider variation of practice by teachers than do questionnaires, but with correspondingly less certainty of how commonly held are the views or how widespread the practices.

Teachers (and to some extent their schools) appear to fall into three categories.

1. Confident, and usually experienced, teachers. They found the new arrangements straightforward and appreciated the opportunity to do less testing of a bureaucratic kind; they used the freedom and flexibility of the new arrangements to reduce their workload (or the intensity of it). For these teachers the new arrangements represented an opportunity to develop and consolidate their existing professional practice.
2. Less confident teachers, often less experienced (though not always). These teachers took longer to see their way through the new arrangements. They benefited from, and appreciated, LEA support / exemplar material, supplementary guidance etc. They were likely to do more work than really necessary (just to be on the safe side) in terms of testing / record keeping / portfolios of children's work. In the end, they were largely successful and would be more efficient in the future because of what they had learned this year. Their workload will have increased this year, but this is most likely to be transitional as their new professional practice becomes established.
3. Teachers who misunderstood the new arrangements or whose existing assessment practice was questionable. These appear to be small in number. They were unable or unwilling to see the new arrangements as a development of existing practice and so found themselves professionally challenged. They tended towards a defensive cynicism: if the National Curriculum tests remain part of the process, they will always

be taken more seriously than any teacher assessment, so ‘what’s the point of doing Teacher Assessment?’ They tended to find the new arrangements meant an increase in workload because they were being required to be more evidence-based and coherent in their Teacher Assessments than was their custom, as well as run the National Curriculum assessments. For these few, this means a structural change in working practice.

It is impossible to be precise about the proportions of the profession that would fall into these three categories. Interviews suggested strongly that teachers were positive about the new arrangements and had made sensible and successful attempts to implement them. In discussion groups there tended to be support for the rhetorical posture of teachers being a much abused profession, and that Teacher Assessment will never be taken seriously by the powers that be. Most, however, also claimed that they took the procedures seriously, found them productive and welcomed the prospective changes. Tentatively, one might suggest a majority showing confidence and success, a smaller number being less confident but still successful and a smaller number than this showing some professional misunderstanding.

Testing and marking

Interviews and focus group discussions, as well as questionnaire comments suggest that the majority of teachers saw great advantage in not having to administer a reading test to all pupils, since the teachers usually felt that they already knew where pupils were from existing classroom practice. They also saw that not every pupil needed to do a maths task as a precursor to a maths test. Some also remarked that there was less need to practise for tests. Escape from the OMR forms was much appreciated.

well yes I didn't have to use the maths tasks at all because all of my children got level 2 or above so I just didn't have to do it... certainly for reading was the biggest difference... that just takes forever doing individual tests with reading... when you do it anyhow I mean I do it on a regular basis... I know how my children read so to have to do it as a test was just too much....

(teacher)

It was much easier not having to complete the LEA - OMR mark sheets

(headteacher)

Less testing also implied less marking, which was another positive. The flexibility meant that both testing and marking could be spread over a longer time and so more easily fitted in.

A smaller number of teachers were more likely to have administered the full National Curriculum test procedure or only to have deviated from it marginally, perhaps for their own peace of mind and as a safeguard against later questions of accountability.

This year we really ran both systems (other than the reporting) so the workload for class teachers was huge. Next year we will reduce the emphasis on the test & preparation and so reduce the workload.

(headteacher)

Evaluation of the trial assessment arrangements for key stage 1

Teachers this year carried out more tasks and tests than were necessary, certainly more than the guidance suggested. However, I feel that this was more a lack of confidence in their own judgements.

(LEA coordinator)

A small group (very small, it is to be hoped) admitted to unusual practices. In one case a teacher said she ran the National Curriculum tests twice: the 2003 test, the results of which would form her Teacher Assessment, and the 2004 test as the 'real thing'. She complained about the extra work involved.

...well the school that we are at... the head was very against assessing the children at all so we didn't have any formal assessment... I mean we do assessments about once a term for writing and that's all we have... we don't have any maths assessment or anything... so when we were suddenly told that we needed teacher assessment... we'd a bit of a problem because we hadn't any official teacher assessment... although we knew where the children were...

...you know... we had the books and things but we didn't have anything official... you know... although we all knew where the children were... so then we suddenly had to think well... what have we actually got? You know... so the teacher that was in charge of Year 2 and the SATS at that time... we did the 2003 SATS as a teacher assessment and then they did their SATS as their SATS... so they ended up doing too much... so it was a bit of a nightmare from that point of view... but hopefully those days have gone.

(teacher)

The last sentiment above led to the revelation of a misunderstanding shared by a number of teachers in this discussion group that in 2005 there would be no National Curriculum tests at all: this year's trial, it was supposed, had been a short transition to a state of assessment entirely by teacher judgement. There was considerable resentment when this was clarified.

Teacher assessment and moderation

The efforts devoted to teacher assessment vary a little according to whether a teacher expects to attend moderation, with an unsurprising shift towards greater record keeping for those anticipating moderation (Table 3.5). Writing and maths show similar patterns. All of those who kept one piece of work per week were in moderated schools.

Table 3.5 – Teachers questionnaire (2) – Extent of work kept

5. To what extent did you keep examples of children's work?		Didn't	One piece per week	Several pieces in each subject	Total
Writing	not moderated	2 (7)	0 (0)	15 (56)	17 (63)
	moderated	2 (4)	5 (9)	42 (76)	49 (89)
Maths	not moderated	4 (15)	0 (0)	14 (52)	18 (67)
	moderated	4 (7)	2 (4)	37 (67)	43 (78)

The totals in the table above are less than 100% as several teachers gave no response or reported having done 'other things' - the most common of which was the use of pupils' exercise books to support the assessment.

Evaluation of the trial assessment arrangements for key stage 1

In commenting on the amount of evidence kept and testing undertaken, it was common to link the quantity both to teacher confidence and to moderation. In relation to confidence, it was observed that the least confident collected up far more evidence and undertook more tests and tasks than the more confident. This was acknowledged by teachers themselves, and was felt to be a reasonable response – a way of becoming more sure.

The effect of moderation on the amount of evidence, including tests and tasks, was generally to increase it, in that teachers from non-moderated schools claimed to have on average kept less evidence than their moderated colleagues, but the degree of the effect did seem to vary across LEAs, and to be related to the moderation systems employed and the clarity of communication with the schools. For example, where teachers could select which children's work would be used, less evidence overall was likely to be collected; where the LEA specified which children's work would be used for moderation (after submission of results by the teacher), the collection of more evidence overall was likely.

The largest effects were where the teachers were not confident and the schools were not clear what was needed for the moderation process. Under these circumstances they were inclined to keep everything they could think of, and not necessarily the most appropriate things.

...a lot of schools when the moderator arrived had all the test and task outcomes there and that was it... they really hadn't twigged that because they were reporting the teacher assessments that's what we would be moderating despite the letter saying moderator might see this or might see that and you know whatever you've got available....

(LEA coordinator)

...we knew quite early on we were going to be moderated and we did get really anxious about it and we were so anxious to do the right thing that I think we probably did more than we needed to and it was quite hard work...

(teacher)

Some teachers were perplexed at the prospect of new arrangements and struggled to make sense of the advance guidance. This caused some stress.

... at the beginning... we felt as though there wasn't that much information about what to do... and we were really worried... and with hindsight we probably didn't need to be but I felt that we were in the dark and I did phone [the LEA coordinator] up once... and I felt I must be being really thick... but I've read this leaflet and it's just not sinking in... and I was really worried...

...yes... we've had exactly the same experience...

...yes... we felt exactly the same...

...but that's probably a pilot year issue because another year I'd probably have hindsight and I'd know more what to expect...

(teachers in discussion)

Another factor in the amount of evidence generated and retained relates to the relative priority being given by the individual teacher (and the school) to obtaining a correct judgement on each child. Where this is paramount, the teacher will do more to be more sure – and this is

Evaluation of the trial assessment arrangements for key stage 1

almost without limit. Where the teacher is trying to balance the need for accuracy in assessment with the need to teach the child, to enhance his or her development, a compromise is reached.

The overwhelming response was that teachers found moderation a positive experience. Many said they learned a lot in terms of how to make accurate judgements. Others gained reassurance.

...and there was a couple of things on there that we picked up that we perhaps weren't doing... but she said it wasn't... that's nothing as a detrimental comment... that was something that don't worry... loads of schools are not doing these... but we came back and put those into practice... but the moderator particularly... she thought she would be with us all day she said... I think she'd gone by eleven... she said I've seen absolutely everything I need to see... your school is fine... so we were always fairly confident that what we were assessing...

(teacher)

Only a few talked of it in workload terms - admitting to feeling that preparation for it was a burden; and these considered that they might have over-prepared for it:

This year we were too conscientious.

(teacher)

[expect increase in workload] in order to do it properly

(teacher)

...the only thing that because we did it as a whole staff... we chose some pieces of writing prior to the moderation up here and like from reception to year 6 we all sat round in groups and we marked it... and that was quite interesting as a moderation process for school... but it does take a lot of time to mark in depth although it is very interesting isn't it that's the only thing that I felt that maybe this year it was extra to your normal teaching time really because it's so rigorous...

(teacher)

School workload

Teachers and headteachers were asked about the impact on the workload in their schools generally, including teaching assistants. Heads were largely of the opinion that the assessments involved a considerable quantity of work for schools, being almost equally divided between it being 'a reasonable amount' and 'a large amount' (Table 3.6).

Table 3.6 - Headteachers questionnaire (2) – Amount of work in doing the assessments

1. How much work was required in doing the assessments? (n = 63)	A small amount	A reasonable amount	A large amount	Too much
	1 (2)	31 (49)	29 (46)	2 (3)

Similar results obtain for those moderated and those not moderated.

Evaluation of the trial assessment arrangements for key stage 1

In comparing the workload with last year, however, heads were firmly of the opinion that it was ‘about the same’ (Table 3.7), implying that if there is a workload issue it is about the burden of assessment in general, and not about the new arrangements *per se*.

Table 3.7 - Headteachers questionnaire (2) – Workload comparisons for management

4. Compare the workload experienced by you/your management team this year with that of last year. (n = 63)	A lot less	A little less	About the same	A little more	A lot more	No answer
	2 (3)	4 (6)	44 (70)	12 (19)	0 (0)	1 (2)

The results are similar for those moderated and those not moderated.

Teachers were asked about changes in the workload of teaching assistants as a result of the assessment arrangements. They were of the firm opinion that it was ‘about the same’ (Table 3.8).

Table 3.8 – Teacher questionnaire (2) – Workload comparisons for teaching assistants

4. Compare the workload experienced by teaching assistants this year against last year. (n = 88)	A lot less	A little less	About the same	A little more	A lot more	Don’t know	No answer
	9 (10)	3 (3)	46 (52)	8 (9)	1 (1)	13 (15)	8 (9)

LEA perceptions

LEA personnel, along with heads and teachers, overwhelmingly welcomed the new arrangements. They often saw many potential educational benefits and felt that any modest change in workload would be a fair price to pay for this.

As hoped, the new arrangements have become (or have the potential to become!) part of quality of on-going teacher assessment for learning. Still some reservations regarding time-commitment for staff though this may become less onerous if allowed, through time, to become a part of the school culture.

(LEA coordinator)

LEA coordinators also considered there to be minimal effects on the workload of headteacher / management team (Table 3.9), with a faint hint that it might be easier - perhaps reflecting a medium to long term view rather than a short-term one.

Table 3.9 - LEA coordinator questionnaire (2) – Effect on management workload

4. To what extent did the trial affect the work of the headteacher / management team? (n = 22)	Made it much harder	Made it a little harder	No change	Made it a little easier	Made it a lot easier	Don’t know	No answer
	0 (0)	1 (5)	14 (64)	3 (14)	0 (0)	3 (14)	1 (5)

In interview, some identified transitional effects:

Any change to existing systems is bound to demand some level of additional work in the transition stage. Some modifications to data systems & reporting arrangements will

Evaluation of the trial assessment arrangements for key stage 1

have impacted on the workload of heads. The potential challenge is the development of whole school assessment systems which is a whole school improvement issue that impacts on leadership and management at all levels.

(LEA coordinator)

3.1.3 The impact of the trial on workload in LEAs

Training and moderation effects

Overwhelmingly, LEA coordinators felt that training and moderation meant ‘a little more work’ for them and their colleagues this year, compared to last year (Table 3.10).

Table 3.10 - LEA coordinator questionnaire (2) – Comparisons of training and moderation workload

1. How did the training and moderation workload for LEA personnel compare with last year? (n = 22)	Considerably more	A little more	About the same	A bit less	Considerably less
	3 (14)	14 (64)	2 (9)	3 (14)	0 (0)

...I think it's changed the focus... it probably is slightly more.... yes... I think that the bigger workload is just because it's something that is new and you have to just re-think what you're doing... I don't think it will be any less work... not for an LEA... I think it just re-focuses it...

(LEA coordinator)

In the light of the role of LEA personnel in supporting and leading assessment in schools, it is perhaps not surprising that the increase in workload, from the LEA point of view, is not necessarily seen in negative terms:

...yes, I think it's possibly a bit more intensive... but then it's also more useful in a way because we've been able to go in and talk to teachers... and talk to headteachers about... you know... how they are doing their ongoing assessments and it's also revealed some other things so that... for example... when we were looking at if we say maths... and looking at children's work across the year and I did some moderation visits with a maths consultant and they said well... this is interesting there is not enough evidence here really of problem solving approaches or maths investigations... and so they were able to get additional information which they'll use when they're working with schools in the future which wouldn't have been revealed just from looking at a test result...

(LEA coordinator)

In one LEA, the moderators (including those who had previously been auditors) felt the work was now much more interesting for them but that considerably more time was needed than for the previous process of auditing tasks/tests. This gave rise to a tension between the time needed to do a good moderation (estimated by a number of LEA personnel as one day for each

school) and the desirability of moderating in as many schools as possible. Because of this, concerns were expressed regarding the funding for moderation.

Data collection

The picture for data collection, however, contrasts sharply with that for training and moderation. More than half of those asked said that data collection had caused considerably more work (Table 3.11).

Table 3.11 - LEA coordinator questionnaire (2) – Data collection workload

2. How did the data collection and analysis workload for LEA personnel compare with last year? (n = 22)	Considerably more	A little more	About the same	A bit less	Considerably less	No answer
	13 (59)	4 (18)	1 (5)	2 (9)	1 (5)	1 (5)

...any new thing is probably an increase in workload with... I would say that the collection of data has been the most difficult thing... and potentially the software could grind everything to a halt... it nearly did with the trial really at some points when people found there were no reports for the assessment... so from that point of view I think our... well not necessarily... well our workload... yes... with our IT team has increased I think...

(LEA coordinator)

However, the data collection problems do not, in the main, appear to be directly related to the trial.

3.2 Management implications of the new arrangements for schools and LEAs

This section describes the management implications of the new arrangements other than those directly concerned with workload.

The section is divided into four parts:

3.2.1 Issues preceding the trial

3.2.2 Issues following the trial

3.2.3 Overall impact on the new arrangements on school management

3.2.4 Impact of the new arrangements of management of assessment within LEAs

3.2.1 Issues preceding the trial

In this sub-section the questionnaire data that preceded (or occurred early in) the trial period is examined to identify possible areas that were thought might have implications for management. This represents the concerns that different stakeholders in the trial LEAs had about the trial in relation to management, and may reflect the concerns of those stakeholders in the remainder of the country if the system is adopted nationally.

Evaluation of the trial assessment arrangements for key stage 1

There were opportunities to express in their own words views related to management impact in the headteacher, LEA assessment coordinator and governor questionnaires. The governor questionnaire contained the open question “What do you see as the greatest challenge to management?”, the first headteacher questionnaire asked: “What do you see as the greatest challenge to management?” and the first LEA coordinator questionnaire contained the question “What do you see as the greatest challenge to school management in the new assessment arrangements?”.

The following summarises the varied responses made.

Ensuring rigour and accuracy

A shared concern for governors, LEA coordinators and headteachers related to the challenge to schools of ensuring their assessment procedures are rigorous and consistent, and that their results are accurate. Governors of all kinds (LEA representatives, parent governors, co-opted governors, staff representatives and so on) noted this as an issue of comparisons through time, and across schools. It was common to see the concern about accuracy related to other schools rather than the respondents’ own. A foundation governor suggested that there is “*no objective way of ensuring that other schools are not marking leniently to obtain better results*”, but many LEA coordinators, heads and teacher governors emphasised the importance of moderation to the issue. Governors of all kinds seemed to feel that training for teachers is the most important factor. This can also be seen as relating to the challenge of building confidence in teacher assessments among the recipients and users of the results, including parents, government agencies and school inspectors, against the background of many people’s faith in the ‘objectivity’ of the test results arising within the separate tests and Teacher Assessment system.

Teacher confidence and morale

Building confidence also relates to teachers’ confidence in their own assessments, and this was another concern of several headteachers, LEA coordinators and teacher governors. Again this occurs against the background of presumed reliance on task and test results. A number of teacher governors in particular spoke of the pressure on teachers to ‘get it right’, and morale was highlighted as important both by headteachers and community / co-opted governors.

It may be noted that in order to ensure consistency of teacher assessment in the new arrangements and build confidence about teacher assessments among others, most LEAs have expanded their moderation procedures to look at classwork as well as the national assessments. Yet this may lead some teachers to feel that their judgements are not trusted, and increase the challenge of building teachers’ own confidence in their judgements.

Quantity of evidence

A further concern expressed by LEA coordinators and headteachers anticipated that teachers would feel the need to justify their assessments by keeping more evidence or by doing more testing than was formally required.

Communication with parents

Many governors anticipated the need to justify assessments, but tended to link this to the importance of satisfying parents’ needs. Expected issues in relation to parents did seem to vary across kinds of governor. Although all kinds of governor highlighted the need for effective communication, LEA governors had concerns about proving that the school has been of benefit to their children; teacher governors referred to helping parents to value the

Evaluation of the trial assessment arrangements for key stage 1

assessment system, including explaining why test results are not being reported; foundation governors worried about parental disagreement with results; and parent governors pointed out that each parent is concerned about their own child, rather than the system as a whole.

Utilising the flexibility

There were concerns expressed among some LEA coordinators in relation to the quality of decision-making in some schools about making use of the flexibility offered by the new system. Others wondered about the capacity of all schools to handle the pressure to maximise results, in the context of some of the opportunities offered by the new system.

Other issues

Specialist LEA coordinators identified some specific potential problems: with aligning the key stage 1 assessment processes with those for the Foundation Stage; with ensuring that teacher assessment covers the full range of the programme of study in English (given the narrowness of any particular task or test); and with teachers being effective in recognising the most useful evidence to keep in mathematics, to avoid keeping examples of work in every aspect of the curriculum, together with an expected difficulty in distinguishing level 2B from 2A and 2C in mathematics classwork for the first time, given that previously the distinction was based on a score on the tests.

Governors identified a range of possible issues for management. Some worried about whether work progress would be sustained over a potentially longer assessment 'season', or about the assessment overtaking teaching, or felt that maintaining enthusiasm might be difficult. Others were concerned about trust between teachers. Parent governors in particular seemed concerned about the possibility of teachers over-working children to get better results.

Many headteachers highlighted finance, and time issues, especially about offering 'quality' time for assessment. Some head teachers anticipated difficulties with operating the system in mixed-age classes.

3.2.2 Issues following the trial

In this section, data from questionnaires, interviews and focus groups following the trial are used to examine the areas identified in 3.2.1 to consider whether, in the light of the experience of doing the trial, the concerns that preceded the trial persisted or were assuaged.

The observations in the section are based on feedback from questionnaires to teachers, headteachers, LEA assessment coordinators and parents, and on interviews with teachers and LEA assessment coordinators, and focus groups with groups of teachers and parents.

In relation to the overall degree of challenge of the management of the trial arrangements, a coding of the responses given in interviews and focus groups showed that a large majority of respondents found the arrangements to be straightforward.

The exceptions to this all seemed to relate to problems in getting the data into the correct form within the software, rather than the assessment arrangements themselves.

The effect of the experience on the issues and concerns anticipated in the questionnaire responses from before the trial, and described above, will be organised under headings related to the main themes of those concerns.

Ensuring rigour and accuracy

Ensuring that assessment procedures are adequate to the challenge presented by the shift to a greater emphasis on teacher assessments is of course one of the main issues in the trial arrangements, and was recognised as such in the questionnaire responses from the beginning of the trial. The need for assessment moderation, as the main mechanism by which consistency and rigour can be achieved, was identified by both headteachers and LEA coordinators as the most important challenge in the trial.

The degree of confidence expressed by headteachers and LEA coordinators in the accuracy of the assessments actually arrived at in the trial (see section 2.3.1) suggests that most school managers succeeded in doing what was necessary to bring about the necessary quality in assessments.

Comments in questionnaires reinforced the overall picture of confidence, but also pointed up some of the specific obstacles to achieving consistency across all contexts. For example, in schools where there is only one Year 2 teacher, and most particularly when that teacher is new to Year 2, there is an enhanced need for some kind of whole-school approach, and / or networking across schools, to prevent that teacher from being isolated in her / his judgements. Another common observation was that the schools with well established assessment procedures were much better placed to deliver the necessary rigour, with the others much more dependent on external moderation procedures.

Another notable cluster of comments speculated about management challenges in the future. Several LEA coordinators suspect many school managers are not yet tuned in to their responsibilities in making the assessments robust, and that when this aspect comes through more (i.e. more than it tends to do in a ‘trial’) it will present greater challenges.

Teacher confidence and staff morale

A further widespread concern before the trial arose from a presumption that many teachers had become dependent on the test and task results, and would struggle to find the necessary confidence in their own judgements. This was expected to challenge morale.

Teachers’ confidence in their own judgements was addressed directly in the second questionnaire (question 16): How confident were you in the judgements you had to make when balancing tests, tasks and children’s work to produce one overall attainment level for each child (requested separately for reading, writing, mathematics). The results are shown in Table 3.12:

Table 3.12 – Teacher questionnaire (2) - Confidence of teachers in producing an overall attainment level

(n = 88)	Very confident	Quite confident	Not very confident	Not at all confident	Don’t know	No answer	Median/mode
Reading	43 (49)	38 (43)	5(6)		1 (1)	1 (1)	Very confident
Writing	41 (47)	38 (43)	6 (7)	1 (1)	1 (1)	1 (1)	Quite/very
Maths	42 (48)	34 (39)	9 (10)	1 (1)	1 (1)	1 (1)	Quite/very

The impressions of LEA assessment coordinators were also sought (question 6): How confident were teachers about the accuracy of the overall attainment levels produced?

Table 3.13– LEA coordinator questionnaire (2) - Views on teacher confidence

(n = 22)	Very confident	Quite confident	Not very confident	Not at all confident	Don't know	Median/mode
Overall	6 (27)	16 (73)	0 (0)	0 (0)	0 (0)	Quite confident

The impressions of LEA coordinators reinforced the outcomes from the direct questioning of teachers, in suggesting that there is a general confidence among teachers about making these judgements. Similarly, an analysis of views about teacher confidence expressed in the interviews and focus groups suggests that almost all teachers felt confident, although four of the LEA coordinators who were interviewed reported working with teachers who they felt needed their confidence boosting. An analysis of the interviews, to reflect the degree to which teachers were felt to be happy with the increased responsibility because of the value that it offered, showed that a very high proportion were happy, with the remainder ambivalent and none unhappy.

The variability in the confidence of teachers was evident in the discursive questionnaire responses. As might be expected, the more experienced teachers (both generally, and in the specific Year 2 context) were more confident than the inexperienced, and in many schools there had been support of the inexperienced by the experienced, which was always felt to be helpful. Moderation, too, was usually felt to be a bolster to confidence, at least when the moderator's judgements largely agreed with those of the teacher, and the tests and tasks also made a positive contribution to the confidence of some teachers when the results confirmed the teacher's judgements. Managing examples of 'conflict', and retaining teachers' confidence when moderators disagree with teachers, or where the test and task results bring surprising outcomes, is a specific challenge to management in the new arrangements.

In relation to morale the evidence was more 'between the lines', although several teachers did express their appreciation of being trusted, and having their professionalism restored, and some LEA coordinators observed that teachers self-esteem had improved in many cases. In line with that, the impression from both questionnaire returns and interviews was that, following a tentative start in some contexts, teachers overwhelmingly responded to the challenge positively, and morale was enhanced rather than depressed by the new arrangements.

I felt that it was a lot better that our professional judgement was valued in that if a child had missed a level by one or two marks they didn't have to prove it in another paper... you could say well I know they are that level and that's what stood....

(teacher)

The confidence of some Year 2 teachers to make assessment judgements was low at first, but in most contexts grew through the experience of implementing the new arrangements, and is generally expected to grow further as the arrangements become 'bedded down'. In other words a 'period of adjustment' should be anticipated, which for some schools may run to more than one year.

It should, however, be noted that where schools had well established assessment systems, and staff were confident in them and in their own judgements, then the period of adjustment was much shorter. The flexibility within the new system did enable some substantial reductions in the workload associated with test and task administration, and this was helpful to the sense of 'right direction', and to the speed with which teachers' confidence in their judgements became established. These schools did not need to change much to have a sense of progress and comfort with the new requirements. Other schools, in which the assessment practices are less

Evaluation of the trial assessment arrangements for key stage 1

well developed, need to change much more before the benefits of the new arrangements become evident to them. Appropriate and targeted support for such schools seems important.

The need for support seems likely to be greater for smaller schools, those with only one Year 2 teacher. Although many such schools have integrated approaches to assessment, so that the Year 2 teacher is operating cooperatively with staff in Year 1 and Year 3 and beyond, the 'natural' opportunity to discuss levels and evidence with another Year 2 teacher is not available in the school itself, so the proportion of teachers who are in need of support from outside the school is greater. That should not be taken to presume that all teachers in larger schools have helpful cooperative opportunities with other teachers, of course.

The accumulated evidence on confidence in the trial suggests that the challenge to management presented by teacher confidence concerns individual cases rather than the profession as a whole. Overall the message from the trial is that the widespread concern about confidence does not represent a real problem. Whatever initial uncertainties there are will, in most cases, soon go, and the confidence of all teachers will grow quickly with experience. However, it is to be expected that a number of teachers will need targeted support. The impact of confidence on accuracy was also widely noted.

Quantity of evidence

Prior to the trial, both headteachers and LEA coordinators expressed a concern that teachers would keep more evidence and / or conduct more tests and tasks than they needed to, which was a management concern.

The section on workload (see section 3.1) suggests that many teachers did collect more evidence than others, but that this was usually felt by the individual teacher concerned to be necessary for them to be confident in their judgements, rather than because it was a requirement.

Both teachers and managers often felt that keeping too much evidence in the first year would be 'inevitable', but that most teachers would learn from experience which evidence they really needed and which they could do without, leading through time to a reduction. It was also often said that whole-school approaches for assessment could help teachers to become clearer about what was needed and what was not. Some teachers were hoping that there could also be specific guidance from LEAs about what would be sufficient. (LEAs tended in their training to focus on the amount of evidence required for moderation, rather than what schools would need to make sound assessment judgements).

Communication with parents

The success of information transfer to parents may be judged from responses to the questionnaire to parents, which was distributed by schools after the Teacher Assessment levels for each child had been sent home by the school.

Table 3.14 - Parent questionnaire – Receiving of Assessments

	Yes	No	Don't know	No answer
1. Have you received the Teacher Assessment levels for your child this year?	536 (94)	20 (4)	13 (2)	1 (0)

Table 3.15 - Parent questionnaire – Teacher Assessment information

2. Have you had enough information about the Teacher Assessment levels?	Yes	No	Don't know	No answer
	393 (69)	130 (23)	40 (7)	6 (1)

A large majority of parents felt that they had received enough information about the Teacher Assessment levels.

Table 3.16 - Parent questionnaire – Information sources for Assessments

3. Where did you get most of the information about the Teacher Assessments from? ³	Speaking to the class teacher	Speaking to other parents	Letters from the school	Government publications	Meeting at school	Other/none
	245 (43)	53 (9)	293 (51)	16 (3)	26 (5)	41 (7)

Table 3.17 - Parent questionnaire – Parental understanding of the Assessments

4. Were the Teacher Assessment levels easy to understand?	Yes	No	Don't know	No answer
	458 (80)	70 (12)	28 (5)	14 (3)

Again a majority of parents found the Teacher Assessment levels easy to understand, but there were a minority of parents agreeing with sentiments like these from one parent:

I agree with SATs/TA although I would have liked to have received more information explaining them in an easy to understand information leaflet.

(parent)

In other words, there seems to have been substantial (though not total) success. In that respect, however, the new arrangements may not represent particular difficulties over and above any other new arrangements – the challenges of getting notes home or attendance at meetings, together with being understood as intended are not specific to these assessment arrangements.

Utilising the flexibility

Before the trial, some respondents, especially at LEA level, suggested that decision making in utilising the proffered flexibility in the arrangements might be poor in at least some cases. The evidence of the trial itself suggested that most teachers did take up the opportunity in some way. Many teachers expressed an appreciation of the new flexibility, firstly in being able to administer fewer tests and tasks and secondly in being able to choose which test to administer and when. Despite misgivings about using the 2003 tests from some teachers, significant numbers of teachers did use the 2003 material (Table 3.18).

³ Multiple selections were allowed here

Table 3.18 – Teacher questionnaire (2) - Uses of different tests and tasks in the 2004 trial

Reading	Test/task	2003	2004	Both
	Level 1 task	27 (31)	31 (35)	1 (1)
	Level 2 task	21 (24)	34 (39)	3 (3)
	Level 2 test	23 (26)	46 (52)	9 (10)
	Level 3 test	25 (28)	53 (60)	8 (9)

Writing	Test/task	2003	2004	Both
	Short task	40 (45)	44 (50)	3 (3)
	Long task	39 (44)	45 (51)	4 (5)
	Spelling test	25 (28)	60 (68)	2 (2)

Maths	Test/task	2000	2001	2002	Several
	Level 1 task	3 (3)	15 (17)	33 (38)	1 (1)
		2003	2004	Both	
	Level 2 test	28 (32)	53 (60)	6 (7)	
	Level 3 test	25 (28)	54 (61)	6 (7)	

The analysis of the views expressed in interviews and focus groups also suggests that most teachers and LEA coordinators are happy with the opportunities that the flexibility affords, with a large majority positive about the flexibility of timing, and about the choice of tasks and tests.

However, some schools did not find it easy to make the decisions, and would have welcomed guidance about criteria on which such choices might be made.

A particular response to the offered flexibility which represents another issue is that of a few schools remaining wedded to an approach to teaching that is test oriented. These are schools that tended previously not to do independent teacher assessment, but to base the reported Teacher Assessment on test and task outcomes, and which have chosen to continue with this approach within the trial. They seemed to use the flexibility to continue exactly as before, suggesting not only that it is easier not to change than it is to change, but also that the approach brings the better results in terms of pupil attainment. Given the self-selecting voluntary nature of the schools in the trial, it might be expected that the number of schools across the whole country who would take this view could be higher than in the trial.

Other issues

The grading of mathematics within a level - that is deciding between 2A, 2B, and 2C - was expected to cause difficulties and, according to LEA coordinators, did for a substantial minority of teachers. This may merely reflect lack of experience, given the previous approach of differentiating by marks in the test, or it could arise from the composite nature of the mathematics curriculum, so that the greater expectation is of children 'doing level X work' than of 'doing their work at level X'. Several teachers, reflecting on this issue, hoped that the guidance material could be supplemented by either 'best-fit' descriptors at the sub-levels, or a detailed breakdown of curriculum elements into sub-levels.

The ability of schools to handle the pressure to 'shape' results, which had been a concern of some respondents before the trial, was also neither realised in the trial nor removed by it. However, several teachers and LEA coordinators expressed the view that it was easier to resist such pressure in a trial than it will be over time.

Evaluation of the trial assessment arrangements for key stage 1

Before the trial, some headteachers had highlighted time and money issues, and while the trial had been conducted pragmatically, with the prevailing conditions coped with, there were several suggestions after it that more money, for example of supply cover to release teachers, would improve the quality of what is done.

Other concerns that had arisen before the trial but which did not seem to have been an issue in the trial itself included: mixed-age classes; the linkage with Foundation Stage profiling; the range of English assessment; the length of the assessment period; trust; parental queries; enthusiasm; children being over-worked; and assessment overtaking teaching.

Summary of Issues following the trial

Anxieties about teacher confidence and the knock-on effect this might have on accuracy of results, in general, proved to be misplaced. Most heads felt that quality was achieved. Related anxieties about teacher morale and dependency on tests were similarly the exception to the rule, though it is acknowledged that a minority of teachers exist who would need support in these areas.

Teachers' responses to the new arrangements were highly positive and the experience was a boost to morale. The majority of parents were likewise very positive about the outcomes. Any failure to achieve total success can be seen as an inevitable consequence of introducing any new arrangements.

The vast majority of teachers have welcomed the flexibility of the new arrangements, though some would appreciate additional guidance in the use of this flexibility when it comes to the choice of tests and tasks.

The grading of mathematics into sub-levels did cause difficulties for a substantial minority of teachers, though this may be resolved as professional experience is accumulated.

3.2.2 Overall impact of the new arrangements on school management

The main major new challenge to management in the new arrangements is in ensuring rigour and accuracy in teacher assessments. Teacher assessment could in the past be left to the discretion of individual teachers with little impact on the school, but its heightened status within the new arrangements places a responsibility on management to ensure that it is done well. The new arrangements also represent an opportunity to improve the accuracy of assessments, by avoiding the distorting effects of one-off tests.

Several elements need to be in place for these changes to be realised in practice, including participation in training about the meaning of levels, developing or maintaining a record keeping system that is helpful to the process, and engaging fully in moderation processes, (and not just at the end of the year and only when the school is in the LEA sample). School, and to an extent LEA, management has to deliver new results in all of these areas.

The small number of schools in the trial that did not have well-established teacher assessment systems in place faced the challenge of establishing them, but the trial was more of a prompt, a forceful reminder of what would be a good idea, than a requirement. It was still possible within the trial arrangements to fulfil the requirements on the basis of test and task results.

Over the longer term there is a high desirability of establishing a whole-school integrated assessment approach, in which Year 2 and Year 6 are no longer isolated as the dual reservoirs

Evaluation of the trial assessment arrangements for key stage 1

of all the important assessments, but this may not and need not be faced in all schools immediately.

Other than this new demand of ensuring rigour and accuracy (and also excepting the data collection aspect) the management implications of the new arrangements represent either new versions of familiar challenges, or matters of specific detail. Managing the new arrangements is by and large more an adjustment than an addition, a matter of formalising slightly the procedures that were already in place for teacher assessment. It is also a matter of applying to this new aspect the common challenges for school management, whatever the issue: creating time for staff to do a good job, and supporting staff, especially the less experienced.

Building teacher confidence in what they do is part of school (and LEA) management responsibility, and its focus on teacher assessment represents an extension of a familiar thrust. This can also be linked to the identified need for management support for teachers in using effectively the flexibility that the new arrangements offer. Supporting teachers in making good choices within teaching is already part of what school (and in a different way LEA) management does, and always involves the development of confidence. Extending that into teacher assessment is only new in some respects – and will be barely different at all if the school is engaged in developing Assessment for Learning.

The same may be said of communication with parents, which is an on-going management issue for all schools.

3.2.3 Impact of the new arrangements on management of assessment within LEAs

The main effect on LEAs of the change of emphasis within the new arrangements has been to lessen their policing role, and to develop more their supportive role. In all LEAs this has been linked to a strategy for assessment development that was already in place, and to investigate new ways to integrate the administration of the statutory assessment and reporting arrangements with that in a way that had been more difficult in the past.

These opportunities have been interpreted strategically differently in different LEAs, with most trial LEAs evaluating their changing role as a result of the experience and likely to change their approach again next year compared to this year. In other words, the LEAs did not feel that they had ‘got it right’ in the trial, but that they had ‘made progress’.

Perhaps the most visible practical change was to moderation, looking at children’s day-to-day work to moderate levelling and teacher assessments, and not just test and task scripts to check on quality of marking. Several LEAs found ways to engage with the opportunity this presented to engage with pupils rather than products as a way of improving validity in the moderation process. LEAs also welcomed the way in which broader based moderation enabled them to understand school processes as a way in to offering more grounded and particular advice, as part of the opportunity that the new arrangements offered to engage more constructively with schools, to get to know them better, which links with LEAs’ broader responsibilities.

LEAs also found that the new arrangements represented an opportunity to promote other long-standing assessment agendas such as Assessment for Learning, and whole-school assessment approaches. The whole-year and whole-school implications of the new system have also brought home the need to involve schools in training on the arrangements from early on in the year, and have raised hopes of integrating the arrangements into LEAs’ strategies for school development and improvement.

Evaluation of the trial assessment arrangements for key stage 1

...it will be more work but actually what it's made us do is joined up our thinking in terms of key stage 1 statutory assessment isn't just about training teachers in this is a test and this is how you mark it according to the mark scheme... it is actually about taking a much broader view about assessment in key stage 1 so next year we're going to be involving year 1 teachers, possible year 3 teachers and we're actually broadening the whole thing... and so yes, it will be more, but it's because we're taking a wider perspective and not making it just about year 2 teachers...

(LEA coordinator)

Greatly improved access to a range of TA judgements will impact well on standards and achievement AFL focus is much stronger. All schools enthusiastic about agreement trialling process (all schools involved) and have requested termly opportunities facilitated by LEA and clusters excellent professional development opportunities.

(LEA coordinator)

Worthwhile as this broader role may be, LEA managers need to remain alert to the importance of what the LEA does to retain the robustness of the assessment process and the consistency of the outcomes.

As one aspect of this, LEAs have some managerial responsibility for the development of teachers in the skills needed to make the assessment arrangements rigorous, through training and moderation. For example, the specific matter of the quantity of evidence that teachers may choose to keep to support their assessment judgements is partly linked to teacher confidence, but is also related to the advice that teachers receive, and to the specific requirements of each LEA in its moderation processes. In that respect, clear and unambiguous guidance from LEAs will go some way to alleviate the issue, although school management will have to acknowledge that some teachers will keep more evidence than is strictly necessary or advised just to be sure that they have enough.

Another specific area that requires management action from LEAs is in the grading of levels, especially within mathematics, which requires further advice and perhaps training.

3.3 Key findings and issues

Workload

- Assessment at key stage 1 presents a heavy burden for schools but the new arrangements have not been seen as adding to that burden significantly.
- Most teachers have reacted positively to the new arrangements. Many have recognised opportunities to reduce their workload.
- Others have been more cautious in the first run of the arrangements and admitted to doing more work than strictly necessary, though this may have brought necessary reassurance. These teachers see the prospect of workload reduction next time around.
- A minority have revealed some professional weakness that will need attention.
- LEA coordinators have tended to see the new arrangements as an opportunity for teacher professional development. In this case any slight increase in workload is seen as the price of good investment for the future.

Evaluation of the trial assessment arrangements for key stage 1

- Data Collection problems this year did increase the workload of LEAs considerably, but this should be short term in that it can be attributed to software problems, rather than the trial itself.

The issues over workload appear to be less significant than was initially supposed. It seems that workload is not seen as an issue when the work done is felt to be helpful towards an end that is desirable. Anything that is seen to be of clear benefit to children, or which is thought of as part of the teaching-learning process, tends to be discounted. 'Workload' seems to be applied only to the actions taken to serve the purposes of others. If teachers are being required to do assessments that they see as unnecessary, or to do anything that is for the benefit of LEA and government bureaucracy rather than for the benefit of children, then the amount of work involved will be monitored and any increase in it will be resented. Even then, however, there can be ameliorating factors. If, for example, a teacher chooses to do more assessments and / or keep more work than they might because they are 'being sure', they will not resent the workload as much – unless this could have been prevented by better advice or clearer guidance.

Within the context of the trial, workload was not a major issue because most of the additional elements were helping to realise a change that was perceived as worthwhile – towards greater professional involvement in the summative assessment process. However, this could have been influenced by the fact of it being a trial, with LEA and schools volunteering and therefore already committed to the purposes of the change. Although it might be expected that most schools will share the trial schools' appreciation of the direction of change, the proportion of those that do not may increase in any roll-out.

Management

This is a complex area not least because of the varied skills of the many people who contribute to the managing of a process such as large scale assessment. Simple, absolutist statements are unlikely to be useful where there is such a degree of contingency. The following statements indicate the findings, along with the issues that arise from them.

- Management concerns prior to the experience of implementing the new arrangements clustered around ensuring rigour and consistency, teacher confidence (with its corollary of over-preparation and excessive collection of evidence) and communication with parents.
- Schools with well-established assessment procedures seemed better placed to accommodate the new arrangements with minimum fuss.
- The management responsibility for ensuring robustness in the teachers' assessments may be felt more deeply in 'roll-out' than it was in the trial.
- Teachers' confidence in their own assessment judgements was seen as vital, and will grow as a result of experience, but cannot be expected to be strong at the start for all teachers.
- Managing 'negative' feedback successfully (when the teachers' judgements are not supported by others or by test and task results) can be a critical point for teacher confidence.

Evaluation of the trial assessment arrangements for key stage 1

- Clear guidance about what evidence to keep to justify assessment judgements is critical, but some teachers still kept more than they were advised to do, especially at first, as a result of insecurity. Over time they will learn what they do not need.
- School managers need to know early in the year about what is involved in order to organise the school resources and timetable appropriately.
- Greater flexibility does not mean total freedom, and school managers need clear guidance about where teachers are free to choose, and where they are not, in order to plan accordingly.

4. Views of stakeholders on other issues

The previous chapters of this report have focussed on a variety of specific areas of interest concerning the impact of the trial arrangements, including robustness of the data, training and moderation, and workload and management implications. This chapter involves a discussion of, and reflection on, the opinions of stakeholders across a broader range of themes in order to complement those already detailed.

The analysis is generally centred around certain key areas that were originally identified as of interest, significance and importance in gauging the degree of success of the new arrangements as judged by those having been part of the trial.

The chapter is split up into the following eight sub-sections:

4.1 The benefits and challenges to schools of the new arrangements

4.2 The classroom experience of Year 2 children

4.3 Parents' views on reporting of the results

4.4 Parents' confidence in teacher assessment

4.5 Linking summative and formative purposes for assessment

4.6 The use of the assessment outcomes

4.7 Comparisons of the new and the old assessment arrangements

4.8 Key findings and issues

4.1 The benefits and challenges to schools of the new arrangements

Views about benefits arising from the new arrangements

The initial questionnaire to head teachers and to LEA coordinators offered five ways in which schools might be expected to benefit, with an invitation to identify those that were expected to occur and to isolate the two that were the most important. A summary of their responses is shown in Table 4.1.

Table 4.1– Heads and LEA coordinator questionnaire (1) - Benefits of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring benefits to your school?	Headteacher (n=81)		LEA coordinator (n = 26)	
	Yes	No answer	Yes	No answer
a. Increased professionalism among teachers	58 (70)	23(28)	26 (100)	0 (0)
b. Increased clarity in outcomes	43 (52)	38 (46)	16 (62)	10 (39)
c. Increased confidence in outcomes	57 (69)	24 (29)	18 (69)	8 (31)
d. Improved communication with parents	22 (27)	59 (71)	5 (19)	21 (81)
e. Fewer parental questions/issues	18 (22)	63 (76)	7 (27)	19 (73)
f. Better experience for the children	61 (74)	20 (24)	26 (100)	0 (0)

Most important		a	b	c	d	e	f
	Heads		34 (42)	13 (16)	26 (35)	0 (0)	4 (5)
LEA		20 (77)	3 (12)	0 (0)	1 (4)	0 (0)	18 (69)

Both headteachers and LEA coordinators highlighted as benefits the expected increased professionalism amongst teachers and a better experience for the children; with headteachers also focusing on increased confidence in outcomes.

In their questionnaire, governors were asked the more open question: “In what way(s) do you think that the new approach to key stage 1 assessment is likely to bring benefits to the school?” Reducing the pressure on children was identified by 59 (36%) governors as an expected benefit, with 31 (19%) noting the enhancement of credibility to teachers and 39 (24%) picking out the likely improved accuracy of judgements that arises from these. Reduced pressures on teachers (21, 13%) and parents (10, 6%) were also noted, with 10 (6%) mentioning that the results will be easier for parents to understand.

All of these views were obtained from questionnaires administered early in the trial, and reflect the anticipated benefits of the new arrangements. After the arrangements had actually been implemented, the interviews with teachers and the teacher focus groups gave opportunities to report experienced benefits.

As will be described in detail below, a reduction in pressure on children was widely reported. There was also a number of teachers who prided themselves that they had kept the pressure to a minimum under the ‘old’ arrangements, so that the new arrangements did not change the experience for the pupils - the difference was just that it was easier for them as teachers to bring this about. No respondents at all said that the pupils’ experience was worse under the new arrangements.

The improvement in professionalism and professional credibility was noted by almost all teachers as a benefit of the new arrangements, and for many this was aligned with an improved experience for them, with less stress often thanks to the flexibility within the system.

Evaluation of the trial assessment arrangements for key stage 1

The questionnaire to LEA coordinators that followed the trial asked, as an open question: “What do you think have been the main benefits to schools of the new assessment arrangements?” The most common response to this (59%) referred to the more educationally relevant assessment practices that the new arrangements permit. The contribution to teachers’ credibility and professionalism was highlighted by 8 (36%) of the respondents, with 8 (36%) also mentioning the effects of the changes on increased accuracy of assessments. Other responses highlighted, as above, the reduced pressure on teachers (23%) or children (14%), the advantages of flexibility (23%), and improved confidence (9%) with one coordinator saying that a real benefit had been that the results are more readily understood by parents.

Views about challenges to the school

Again, the initial questionnaire offered five possible areas for headteachers and LEA assessment coordinators to identify as challenges, with an invitation to note the most important in their view.

Table 4.2 – Heads and LEA coordinator questionnaire (1) - Challenges of the trial arrangements

	Headteacher (n=82)		LEA co-ordinator (n=26)	
	Yes	No answer	Yes	No answer
In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring new challenges for the school?				
a. Greater need to formalise teacher assessments	37 (45)	45 (54)	16 (62)	10 (39)
b. Greater need for transparency	24 (29)	58 (70)	10 (39)	16 (62)
c. Greater need for moderation of assessment	65 (78)	17 (21)	25 (96)	1 (4)
d. Greater need to justify assessment procedures	32 (39)	50 (60)	13 (50)	13 (50)
e. Greater need to justify assessment judgements	48 (58)	34 (41)	15 (58)	11 (42)

Most important		a	b	c	d	e
	Heads		15 (14)	9 (8)	49 (44)	9 (8)
LEA		8 (22)	5 (14)	17 (46)	3 (8)	4 (11)

Overall, a greater need for moderation of assessment seems to have been seen by both headteachers and LEA coordinators as the most important new challenge.

The questionnaire for governors asked the open question: “In what way(s) do you think that the new approach to key stage 1 assessment is likely to bring new challenges for the school?”

Only 5 governor respondents (3%) mentioned ‘moderation’ explicitly. However, the most common response (24%) was to point up the challenge of maintaining consistency and accuracy in the assessments which is essentially the purpose of moderation. Other than this, a common response (15%) was that the challenge was really just adjustment to the changes rather than anything specific, and many governors either did not respond (40%), or said explicitly (7%) that they did not anticipate any new challenges, e.g. “None envisaged as the staff are used to doing thorough assessments”.

After the trial, views about what was found difficult or challenging were sought from teachers in the interviews and focus groups. It seems from these responses that the changes to moderation were less challenging than heads and LEA coordinators had anticipated, with

Evaluation of the trial assessment arrangements for key stage 1

most reporting positively on the experience with LEA moderators (see also section 2.2.2) and those that undertook internal school moderation being very positive about that – although of course what was done as internal moderation varied a great deal and was not done at all in many schools. The one fairly common issue relating to moderation was in respect of collecting evidence to justify assessment judgements, with many teachers collecting what they felt to be too much, in order to be sure of having enough.

Several challenges arising from the new arrangements concerned adjustment, as anticipated by governors, for example in getting used to the flexibility that was offered. Other teachers reported a challenge in getting used to the meaning of levels and sub-levels in relation to everyday work, that is learning how to evaluate work in more detail than had been needed previously, especially in mathematics.

LEA coordinators were asked the following question in the questionnaires: “What aspects of the new arrangements have schools found most challenging?” The most common response to this (41%) concerned teachers learning to rely on their own judgement, and not testing more than necessary. A further 9% of respondents mentioned the effective use of the flexibility within the new arrangements. The issue of how much evidence to collect was highlighted in 9% of the responses, and the challenge of judging levels (23%) especially mathematics levels (18%) was also common. Otherwise, the only common element (23%) was the difficulty surrounding data collection.

4.2 The classroom experience of Year 2 children

The overall feeling in this regard is that the new system has generally improved the experience in school of children in Year 2. One clear piece of evidence of this is in the contrasting views of headteachers on the arrangements in 2002/3 compared to the trial arrangements this year (2003/4):

Table 4.3 – Headteacher questionnaires – The impact of the trial arrangements in the classroom

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer	Medial /Modal	Agreement index ⁴
5. They had a positive effect on teaching and learning in the classroom. (n=83) (2002/3)	22 (27)	40 (48)	16 (19)	1 (1)	1 (1)	3 (4)	Disagree	-56
10. They had a positive effect on teaching and learning in the classroom. (n=63) (2003/4)	1 (3)	12 (19)	31 (49)	10 (16)	6 (10)	3 (5)	Agree	45

There is a clear sense here that in the opinion of the heads the trial arrangements have made a considerable positive impact in the classroom with both their modal and medial opinion changing from Disagree (in the earlier questionnaire) to Agree in the latter. A chi-squared calculation confirms this as a statistically significant effect ($p < 0.001$).

Evaluation of the trial assessment arrangements for key stage 1

Heads and LEA coordinators were also asked (in the first round of questionnaires) about potential benefits of the trial arrangements and these gave the following specific answers, amongst others.

Table 4.4 - LEA coordinator and headteacher questionnaires (1) – The benefits to the school of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring benefits to your school?	LEA coordinator (n=26)	Headteacher (n=83)	
	Yes	Yes	No answer
Better experience for the children	26 (100)	61 (73)	22 (27)

Again, the evidence in Table 4.4 provides a strong endorsement of the view that the trial has improved the children's experience in Year 2.

Other evidence on this subject can be gleaned from some of the teacher interviews and focus groups. Here feelings were more mixed with perhaps the majority of teachers reporting that the new arrangements had not made a large impact on what actually happened in the classroom. This seems to tie in with schools where the KS1 testing regime had remained largely unchanged from last year into this (the trial) year, though it should be pointed out that this does indeed include many schools where the tests are kept very low-key and are not seen as a particularly negative experience in themselves. Consequently the experience in many schools remained similar to that in earlier years, both good and bad.

In contrast, in those schools that reported doing less testing (significantly less in some cases), feelings about the effect of the trial on pupils was far more positive. There was less pressure on the children, with less tests taking place, often with the tests being treated in a less formal manner, and consequently more time for teaching. As one teacher commented there is:

not a full term of tests anymore

(teacher)

There was also the possibility of earlier, more diagnostic testing, that would then inform teaching in the remainder of the year and this was mentioned across a few sessions.

Teacher comments from their second questionnaire concerning the experience of children in the trial were also broadly positive,

The trial made it less stressful for teachers but more importantly less stressful for the children

(teacher)

...I don't think the children suffered for it at all...

(teacher)

..we've always prepared them well in advance and we've never put them under ... test pressure as such ... so I can't really say that the children have felt any less pressure because they've never felt pressure undertaking the test... we've always tried to avoid that and prepare them well in advance...

(teacher)

Evaluation of the trial assessment arrangements for key stage 1

but did include a small number contradicting the majority and bemoaning the continued negative effect of the tests :

Still very stressful for staff and children (due to parents expectations, not staff)

(teacher)

In all the evidence gathered during the evaluation no negative teacher opinions were expressed on the trial's impact on pupils. Therefore, taking the whole set of professionals consulted (teachers, heads and LEA coordinators) together, there are no clear comments that the child's experience has deteriorated as result. The vast majority believe either that the trial has had little if any effect or are positive about its impact.

In parent focus groups there were mixed views and some uncertainty as to how the trial arrangements had affected children. This is largely to be expected as most parents are making an absolute rather than relative judgement on their child's Year 2 experience unless they have had another child going through KS1 in the recent past. The majority opinion overall across the six formal sessions held was that there was still too much testing at such a young age, though in some schools this was handled very well and children did not really know they were being tested at all. There was a minority opinion expressed in one group that testing was part of schooling and that, although unpleasant at this age, it was seen as acceptable.

Where the trial had led to a significant decrease in the amount and/or a lowering of the formality of the testing, parents seemed much happier with the new arrangements. On parent questionnaires, those parents who had experienced KS1 as parents before were generally quite positive about the new arrangements, presumably partly on the basis that the system had helped improve the school experience for their child compared to their recollection of how elder siblings had fared. This sub-group of parents were generally in favour of the new arrangements, though with a considerable number being unsure. Table 4.5 summarises the responses of those parents who had already experienced KS1 assessment in previous years:

Table 4.5 – Parent questionnaire – Comparing the trial arrangements to that in previous years

9b. Are this year's arrangements better? (n = 246)	Yes	No	Don't know	No answer
	122 (50)	40 (16)	79 (29)	5 (2)

Many of those answering "No" to this question appeared to then contradict themselves in their 'Other comments' complaining about too much testing or testing being inappropriate at this age. This suggests that not all parents (and/or schools) had completely understood the full implications of the trial and had not appreciated that, despite tests continuing, the trial did provide for a much more flexible approach and for less testing in absolute terms. It appears a significant number answered No to this question on the basis that the current trial involves testing and is therefore not to be supported. One such parent commented:

I think its unnecessary for them to go through these tests at a young age. 11 years is soon enough!

(parent)

The majority of 'Other comments' by parents were positive towards the trial in general:

Far better system - less pressure on children

Evaluation of the trial assessment arrangements for key stage 1

(parent)

I agree that Teacher assessments are a vital way of assessing a child's learning,

(parent)

... parents felt less stressed I think than in the past... I've had Year 2 for several years and the parents were quite happy about it... and they felt that oh it's a good job they haven't done as many of those tests this year... they were very positive...

(teacher)

However, some were negative about testing at this age:

I do not think that Year 2 children should be tested

(parent)

I mean there is too much pressure put on children at seven years old... what's the knock on effect later on in life you know this is what they're saying like the Year 6 thing there is a new thing isn't there... people are worried about that having an effect on them when it becomes the pressure of 'O' levels and as they lead on into life... it's what effects them when they are younger I think can effect them throughout their lives... if you start putting pressure on them young...

... yes quite often young children can't express as well their anxiety... you know older children can display it in lots of ways but like seven year olds sometimes they find it difficult to say how they are feeling and they go through it and then it comes out in other ways you know stop sleeping stop...

...they don't always understand how they are feeling do they it tends to come out in different ways....

...they don't know what feelings are sometimes do they at seven...

(parents' exchange)

There were a few comments lamenting the perceived undermining of the National Curriculum tests at KS1:

Teacher Assessment is a poor tool

(parent)

As a final observation on the effect of the trial on the children, governors were asked in their questionnaire to describe in which ways the trial arrangements would bring benefits to their school. Approximately 39% gave an answer that included the feeling that there would be less stress or pressure on children under the new system. The following comments are typical:

Pupils will not have so much pressure put on them if there are fewer tests

(governor)

and

Less stressful and fairer on the children

(governor)

Similar sentiments were expressed by governors in the ‘Other comments’ section with many taking the arguments one step further:

Do 7 year olds really need this type of testing? I don't think so

(governor)

4.3 Parents’ views on reporting of the results

Only a few parents felt they were not being given enough information on their child’s progress at the end of Year 2. Only a small minority asked for their child’s test levels as evidenced in the parent questionnaire.

Table 4.6 – Parent questionnaire – The demand for test results

8a. Did you ask for any of your child’s Test levels? (n = 570)	Yes	No	No answer
	82 (14)	475 (83)	13 (2)

When split by LEA, the percentage of parents asking for results did show a very large variation from only 4% in one LEA up to 38% in another. This might indicate that parents’ confidence in teacher assessment varies considerably across LEAs, or possibly between schools within an LEA, but since when looking at LEA-specific data the numbers can get quite small (down to only 20 parents from just four schools in one particular LEA) there could be a sample effect.

Across all the responses given, over 90% of parents said they asked for the test results because they like to have all available information, and only 2.4% because they were not satisfied by the accuracy of the assessments.

Table 4.7 gives the responses by those parents who did not ask for their child’s test results.

Table 4.7 – Parent questionnaire – Reasons why the test results were not asked for

8b. If no, why not? ⁵ (n = 475)	The Teacher Assessment levels gave me enough information	I don’t think the Test Levels are very important	I didn’t know I could request the Test Levels	I didn’t know the Test Levels existed	Other
	196 (41)	47 (10)	216 (46)	45 (10)	36 (8)

Clearly, many parents (56%) were unaware or had perhaps forgotten that they could request the test levels separately from the teacher assessments. That the test results were available from schools is clearly set out in the QCA letter KS1 parent leaflet http://www.qca.org.uk/ages3-14/downloads/KS1_parent_leaflet.rtf though unfortunately the

evaluation provided no specific data on how many schools made this letter available to parents.

Evaluation of the trial assessment arrangements for key stage 1

Of the 41 parents who made a written comment to this question, mostly having answered 'Other' to the first part, 23 (56%) indicated that the test results had or would be provided to them automatically, either in their child's report, or at a parents' evening. One parent said that in light of our questionnaire they would now request them. One other parent did not know what 'test levels' were.

In spite of some parental confusion, Table 4.7 shows that just over 50% of all parents were either happy enough with the Teacher Assessment results, or were not concerned about levels anyway. This does however indicate that the other 50% might have some interest in these results, though there cannot be any degree of certainty about this conclusion without further investigation.

Teachers and headteachers were also asked about parental demand for test results and reported very similar low levels of interest in the test levels across the three main subjects:

Table 4.8 – Teacher questionnaire (2) – Parental demand for test results

15. How many parents specifically asked for their child's test results? (n = 88)	None or almost none	Some	Many	All or almost all	No answer	Median/mode
Reading	67 (76)	12 (14)	2 (2)	1(1)	6 (7)	None or almost none
Writing	67 (76)	12 (14)	2 (2)	1(1)	6 (7)	None or almost none
Maths	65 (74)	13 (15)	3 (3)	1(1)	6 (7)	None or almost none

3 teachers said the results had not yet been sent, and 2 teachers said the test/task results were sent out in the report anyway.

Table 4.9 – Headteacher questionnaire (2) – Parental demand for test results

16. How many parents specifically asked for their child's test/task results? (n = 63)	Few or none	Some	Many	All or almost all	Don't know	No answer	Median/mode
Reading	47 (75)	4 (6)	3 (5)	3 (5)	2 (3)	4 (6)	Few or none
Writing	47 (75)	5 (8)	2 (3)	3 (5)	2 (3)	4 (6)	Few or none
Maths	47 (75)	5 (8)	2 (3)	3 (5)	2 (3)	4 (6)	Few or none

Again, there is a general indication that there was no great desire on the part of the parents for more information than they were given in most schools under the trial's statutory requirements.

This area was also discussed in the parent focus groups, where four out of the six groups were generally judged to be in favour of the new arrangements, with one group being ambivalent complaining about the degree of testing being inappropriate at this age, and one group being slightly negative, feeling that information was being lost to parents since they were not being given the test/task results automatically.

Evaluation of the trial assessment arrangements for key stage 1

*...I think you do need to have the mix between the two [tests and teacher assessments]
...to give you a real general overview of what your child is capable of.*

(parent)

Another group, whilst generally in favour of the arrangements, also wanted the test/task results to be reported automatically to them. The remaining three groups appeared generally content to go along with Teacher Assessment and felt no particular desire to see the test/task results, though one parent did comment that it was nice that they were available if needed.

...it is nice to have the option, because if you're not...I mean I'm really happy with [the teacher but there might be years when you're not really as happy with one teacher as another, so it's nice to have the option...

(parent)

From the questionnaire, see Table 4.10, there was a general feeling from parents that they had been given enough information about the Teacher Assessment levels.

Table 4.10– Parent questionnaire – Teacher Assessment information

2. Have you had enough information about the Teacher Assessment levels? (n = 570)	Yes	No	Don't know	No answer
	393 (69)	130 (23)	40 (7)	7 (1)

There is, however, a significant minority of around 30% of parents who felt that they hadn't been kept well-informed or were unsure.

I would have appreciated a verbal presentation about what was involved and which areas are being tested specifically.

(parent)

As far as understanding the results they were given, there is obviously some way to go before parents feel that they are completely cognisant of the meaning of the results for their child. For example, teachers and headteachers were asked (Table 4.11) about parental understanding of the pre-trial (2002/3) and trial (2003/4) arrangements:

Table 4.11 – Teacher questionnaires – Parental understanding of the assessment arrangements

The assessment arrangements were clear and well understood by parents.	Yes	No	Don't know	No answer
Q8 2002/3 (n = 115)	22 (19)	57 (50)	29 (25)	7 (6)
Q13 2003/4 (n = 88)	23 (26)	24 (27)	39 (44)	2 (2)

The corresponding questions given to headteachers are detailed in Table 4.12.

Table 4.12 – Headteacher questionnaires – Parental understanding of the assessment arrangements

The assessment arrangements were clear and well understood by parents.	Yes	No	Don't know	No answer
Q8 2002/3 (n = 83)	24 (29)	43 (52)	13 (16)	3 (4)
Q12 2003/4 (n = 63)	23 (37)	12 (19)	25 (40)	3 (5)

Evaluation of the trial assessment arrangements for key stage 1

The picture here is a little complicated. Certainly the pre-trial arrangements were generally poorly understood by parents. Focusing on the 'Nos' it can be seen that despite a large reduction from the previous year there remained (at the time of the completion of the questionnaire) some uncertainty about the current trial arrangements. Speculatively, this might have been to some extent overcome towards the end of term as parents actually received their child's results and met with teachers to discuss them.

Chi-squared comparisons of the teacher responses on parental understanding of the two systems (Table 4.11) give a statistically significant result ($p < 0.001$). The headteachers' responses (Table 4.12) give a similar result ($p < 0.001$). This means that there has been an important and significant improvement in the level of understanding in the new arrangements for the parents as viewed by the teachers and heads. As this was one of the main expectations for the trial system - that it would improve parental understanding of assessment at key stage 1 - this result should be welcomed as a clear success of the trial.

A common pattern about information to parents that arose mostly in the parent focus groups was the difficulty parents had in recognising the information that they had in fact received from the school. They would often claim not to have received a form, supposing that it had been lost in transit, or some such, only to acknowledge later some dim recollection of receiving it, without having processed fully the information it contained. The volume of information passing between school and home was often cited as a factor in their difficulty. One parent pointed out also that many parents cannot read well enough to deal with the information that they receive from the school in paper form. Direct face to face meetings were referred to as a preferable method, both for being more memorable and for overcoming issues of literacy, but of course not all parents can or do attend such meetings.

One LEA coordinator in their second questionnaire mentioned that one of the benefits to schools of the trial was that

Results are more understandable by parents

(LEA coordinator)

Heads were asked in both their questionnaires (Table 4.13) to judge how well the assessment arrangements provided meaningful information to parents.

Table 4.13 – Headteacher questionnaires – The assessments as effective feedback to parents

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer	Median / Mode	Agreement index ⁶
6. They gave meaningful information and feedback for parents on their children's progress. (n=83) (2002/3)	12 (15)	42 (51)	25 (30)	2 (2)	0 (0)	2 (3)	Disagree	-33
11. They gave meaningful information and feedback for parents on their children's progress. (n=63) (2003/4)	1 (2)	7 (11)	42 (67)	10 (16)	2 (3)	1 (2)	Agree	71

These two sets of results indicate quite a shift between the two systems with the trial arrangements much more favoured by heads over the previous system in terms of informing parents. Again, a chi-squared calculation confirms this as a statistically significant effect ($p < 0.001$).

4.4 Parents' confidence in teacher assessment

The issue of whether parents trust teacher assessment to provide a fair and reliable judgement on their children's progress at the end of Year 2 is obviously extremely important to the success or otherwise of the trial. The vast majority of those replying to questionnaires seem to do so (Table 4.14).

Table 4.14 – Parent questionnaire – Agreement with child's levels

5. Did you agree with all of the Teacher Assessment levels that your child got? (n = 570)	Yes	No	No answer
	467 (82)	65 (11)	38 (7)

There were many 'Other comments' agreeing with the following parents:

I have great confidence in the teacher (parent)

I trust the teachers' judgement (parent)

with only a very small minority echoing the following sentiment:

teacher's assessment are less objective as they are influenced by the way the child is perceived by the teacher (parent)

Sometimes parents are also governors and amongst the 37 parent governors returning a (governor) questionnaire, 18 gave a comment of which 5 (28%) could be characterised as

Evaluation of the trial assessment arrangements for key stage 1

against testing at this age at all and another 6 (33%) were broadly in favour of the trial arrangements. There were two parent governors whose comments could be perceived as somewhat negative, though both of these were to some extent ambivalent about the trial:

You may find parents getting angry with teachers when they find out that their child, for example is not taking as much testing as their neighbours!

(parent governor)

As a last comment here it should be pointed out that confidence in the accuracy of teachers' judgements was quite high among headteachers and LEA coordinators (see section 2.3.1), but there was a notable feature in the responses from parents. Parents were almost all much more confident in the judgements of the teacher of their own children, and in their own school (that they knew and had come to trust), than they were about the accuracy of the judgements of teachers in other schools (whose motives were mistrusted, and whose professionalism was not recognised).

4.5 Linking summative and formative purposes for assessment

Before the trial, headteachers and LEA coordinators shared an expectation (widely expressed through the first questionnaires) of using the opportunity presented by the new arrangements to link formative assessment (Assessment for Learning) to the summative Teacher Assessments at the end of the key stage. Interviews revealed that this is a strong agenda for LEAs, and in their view is vital to the continuing success of assessment arrangements with a strong teacher assessment element.

... with schools taking on more assessment for learning next year as a priority, it [the trial arrangements rolling out] should be helpful.

(LEA coordinator)

However, the linking of the summative assessment of reported levels to formative assessment that was hoped for by some at the start of the trial did not seem to have been realised in any new ways or in any new contexts. The schools that already had an established teacher assessment system integrated with planning and teaching were able to continue to use it, but the changes in the trial did not seem to represent an impetus to re-organise other assessment approaches to make the assessments more useful to teaching and learning.

Where LEAs made suggestions about how assessment might be approached that integrated the assessment requirements of the trial with assessment for learning, this was little or no different from the advice given previously. However, the experience of the trial does not seem to have dampened the hope that there will be change in this direction, and there were still some respondents expressing the view for the future, saying that there is more potential for integration because summative assessment is not so much 'end-loaded' in the new system. It was also pointed out that linking formative and summative purposes takes time and effort to establish, possibly a longer time period than the trial allowed. The process has begun, but will have to continue.

I think realisation will begin to dawn especially with assessment for learning taking a higher profile. We've always tried to give it a high profile but it now seems to be getting government backing. I think that message will go through to head teachers much more strongly than it has in the past....

(LEA coordinator)

4.6 The use of the assessment outcomes

There is an understandable degree of uncertainty about how any proposed change to end of Year 2 assessment will effect the use of the KS1 results data. For example, the LEA coordinators were asked about this in their first questionnaire and responded as follows:

Table 4.15 – LEA coordinator questionnaire (1) – Change in use of assessment outcomes

12. Does the LEA intend to change its use of the outcomes from key stage 1 assessment in school analyses such as value-added comparisons? (n = 26)	Yes	No	No answer
	4 (15)	15 (58)	7 (27)

Although the majority indicated that no change was intended, many of both the Yes and No respondents then answered the follow-up question:

Table 4.16 – LEA coordinators questionnaire (1) – Actual change in use of assessment outcomes

12b. If yes, please describe how (n = 13)	It will be based on TA	Currently under review
	3 (23)	10 (77)

Their views indicate an uncertainty that is unlikely to be resolved until any change to the KS1 assessment arrangements are fully confirmed.

The views of Junior schools in this regard are obviously important and in their questionnaire gave the following responses:

Table 4.17 – Junior school questionnaire – Change in use of assessment outcomes

3a. In the trial, and possibly from now on, the end of key stage 1 assessment will just be a Teacher Assessment. Is this likely to affect your use of the assessment information? (n = 27)	Yes	No	No answer
	11 (41)	15 (56)	1 (4)

indicating that some changes are anticipated. The responses to the follow-up question were coded as follows:

Table 4.18 – Junior school questionnaire – Reasons for anticipated change in use of assessment outcomes

3b. If yes, please explain why ⁷ (n = 16)	TA possibly more secure - not just a snapshot	TA possibly less secure - not standardised	Level indicators different at KS1 and ks2	More collaboration across KS1/ks2 boundaries	Will continue to use other tests
	1 (7)	4 (27)	2 (13)	4 (27)	5 (33)

This indicates a certain lack of confidence in Teacher Assessment at KS1 and might mean that the amount of initial testing early in Year 3 could increase a little under roll-out. However, the further follow-up (Table 4.19) tends to undermine this argument.

Table 4.19 – Junior school questionnaire – Likelihood of doing internal initial assessments

3c. Specifically, are you more likely or less likely to do your own initial assessments? (n = 27)	More likely	Less likely	No different
	9 (33)	1 (4)	17 (63)

A clear majority indicated that the trial arrangements under roll-out would make no difference. There is, however, a large minority (one third) that do think it more likely that additional assessments will be required under a more teacher assessment-led regime. Views on this might be likely to change over time as teacher assessment becomes the norm at KS1 and confidence in both the process and results grow, for both the Year 2 and Year 3 teachers.

The final follow-up in this series on the Junior school questionnaire (Table 4.20) asked for reasons as to why more testing might take place.

Table 4.20 – Junior school questionnaire – Reasons for doing internal initial assessments

3d. Why are you more likely to do your own initial assessments? (n = 16)	To ensure baseline accurate – concerns about grade inflation or the different meaning of Level 3s at KS1 and KS2	KS1 results need supplementing	Both of these reasons
	10 (63)	5 (31)	1 (6)

In the teacher focus groups this issue was also touched upon by asking “*How have they [First/Infant schools] reacted to the new system?*” and answers to this varied but there did not seem to be a wide amount of reported concern within schools with a Year 3 intake, though it is not clear how many of these schools have a clear understanding of the proposed changes being made at KS1.

...I don't think they've actually registered it [the trial arrangements] yet...

(teacher)

Many Year 2 teachers reported that they were already aware that in Year 3 the children are routinely re-tested due to regression over the summer anyway. This would obviously continue as before regardless of roll-out.

In relation to this, the grading within levels (sub-levels A, B and C) was generally welcomed by Year 2 teachers, and despite the difficulties for some teachers of doing this, especially in mathematics, many others expressed their desire for sub-levels to be introduced for levels 1 and 3, and into the other curriculum areas (science and speaking and listening). This was partly to increase the amount of information that could be made available to teachers in Year 3 and be used to inform decisions about the children (for example, preventing “level 3” being thought of as “well within level 3” rather than “just in level 3”, and so making expectations more realistic), but also to enable quantitative judgements about progress for all children.

The whole-school implications of such a change were acknowledged but welcomed, reflecting the tendency for the trial arrangements to influence schools to adopt whole-school approaches – starting, for example, with agreement trials, moderation meetings, and levelling exercises, and extending into record keeping systems and even reporting to parents.

Evaluation of the trial assessment arrangements for key stage 1

There was no real sense in the responses to the governor questionnaire that there were concerns with this group about the new arrangements having a negative impact on value-added comparisons, PANDAs or generally becoming more problematic as the baseline for key stage comparisons. Only one governor mentioned this issue:

More schools may need to be moderated to check carrying out correctly in order to ensure data remains valid. Otherwise use of data in PANDA etc will become less useful/accurate.

(governor)

4.7 Comparisons of the new and the old assessment arrangements

Teachers

From a comparison of their first and second questionnaires (Table 4.21 and Table 4.22), teachers seem quite clear that the old arrangements were far inferior to the trial ones in terms of their effect on classroom practice and on their usefulness as good reporting tools for parents:

Table 4.21 – Teacher questionnaires – The impact of the trial arrangements in the classroom

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer	Median Mode	Agreement index ⁸
5. They had a positive effect on teaching and learning in the classroom. (n=115) (2002/3)	38 (33)	49 (43)	18 (16)	0 (0)	3 (3)	7 (6)	Disagree	-60
11. They had a positive effect on teaching and learning in the classroom. (n=88) (2003/4)	7 (8)	26 (30)	32 (36)	18 (21)	5 (6)	0 (0)	Agree	19

Table 4.22 – Teacher questionnaires – the assessments as effective feedback to parents

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer	Median / Mode	Agreement index ⁸
6. They gave meaningful information and feedback for parents on their children's progress. (n=115) (2002/3)	17 (15)	58 (51)	27 (24)	2 (2)	4 (4)	7 (6)	Disagree	-40
12. They gave meaningful information and feedback for parents on their children's progress. (n=88) (2003/4)	6 (7)	21 (24)	40 (46)	16 (18)	5 (6)	0 (0)	Agree	33

Evaluation of the trial assessment arrangements for key stage 1

These generally negative views of the old systems and positive views of the new very much echo the headteachers' on the same issues as shown earlier (see Table 4.13). Again, chi-squared calculations confirm these changes in attitude as statistically significant ($p < 0.001$).

Teachers are generally supportive of the move towards teacher assessment that the trial has brought. By implication there is regret that under the pre-trial arrangements teacher assessment was very much overshadowed by testing and the publishing and consequent attention paid to the test results. Little attention was seemingly paid to their teacher assessments under the pre-trial arrangements and this might well have undermined confidence.

One clear indicator of their opinions in this area comes from their very strong support for national roll-out as reported in their second questionnaire:

Table 4.23 -Teacher questionnaire (2)– support for roll-out

Q17 Do you think that the trial arrangements should be made permanent from 2005? (n = 88)	Yes	No	Don't know	No answer
	72 (82)	7 (8)	7 (8)	2 (2)

In the same questionnaire there were many comments similar to these teachers':

Teacher assessment gives a more rounded picture of a child's achievements

(teacher)

I feel the whole process has been very positive all round. I felt I had been given my 'professionalism' back

(teacher)

There were a substantial minority that would like the trial to have gone further:

Feel it was not necessary to carry out assessment tasks, outcomes were extremely similar to my own TA's...I feel this time could have been put to better use

(teacher)

Opinion was more or less evenly divided among teachers about the contribution of the tests and tasks to the process of achieving accuracy in their judgements. Where some felt that the tests and tasks offer a helpful external reference, many others felt that they did not contribute other than as an additional source of information, without deserving the special status they retained. For some this related to the needlessness of another assessment just to confirm what was already known. For others it reflected a belief that the tests and tasks were in any case not as accurate as teacher assessment, because, for example, children do perform differently on different days, so that a 'snapshot' is inherently less reliable than a balanced judgement arrived at through time.

There were also some issues with specific assessments, for example the 2004 writing task, which seemed to some teachers not to reflect the curriculum; and the mathematics tests, where it is possible to gain enough marks to be given level 3 without succeeding on any of the questions that reflect the level 3 curriculum.

There was a very small number of teachers who were not at all happy with how the trial had gone and gave comments indicating perhaps that the flexibility on offer had not been fully

Evaluation of the trial assessment arrangements for key stage 1

grasped or at least implemented, possibly for very good reasons:

I agree that teacher assessment should take priority but I still think it is too much having such a concentrated test time in May

(teacher)

The teacher focus groups were also very supportive of roll-out, with only one of the eight focus groups not fully supportive. Those in favour of roll-out did always provide caveats, generally to do with the importance of getting information out early, concerns about marking of Writing, the need for better guidelines about amounts of moderation evidence and so on. The strength of feeling was made clear in some groups when asked what their reaction would be if roll-out did not take place:

...it would be bad..... very hard to take.....thanks for trying but you've not been very good at Teacher Assessment...

(teachers)

In the one unsupportive focus group there was an apparent confusion about the nature of the trial with the majority of teachers at the session under the misapprehension that the trial was part of a more or less guaranteed transitional process to test-free KS1 assessment from next year under roll-out. For this reason, they did not support roll-out despite being far from happy with the pre-trial assessment arrangements that involved more testing and less flexibility than was on offer under the trial arrangements.

In the teacher interviews the great majority was again in favour of roll-out and where there was some ambivalence in two or three cases, there was either a sense of the trial not living up to expectation (i.e. the assessment regime in the school remaining largely unchanged from previous years) or concerns about future moderation and how much evidence would be required. This teacher worried about moderation and would actually have preferred to rely entirely on teacher assessment at KS1 and for all testing at this key stage to be eliminated.

Headteachers

Heads were also asked their opinion on roll-out and responded very favourably:

Table 4.24 – Headteacher questionnaire (1) – support for roll-out

18. Do you think that the trial arrangements should be made permanent from 2005? (n = 63)	Yes	No	Don't know
	49 (78)	5 (8)	9 (14)

During a roundtable interview with two heads and an LEA coordinator one of the heads relayed his initial feelings on hearing some time ago about the proposed trial arrangements stating that they (the new arrangements) appeared:

...too good to be true...

(headteacher)

and when asked about the possibility of reverting to last years' arrangements if roll-out were not to happen responded:

...disaster would be putting it mildly...

(headteacher)

Evaluation of the trial assessment arrangements for key stage 1

Of course, not all heads are as positive, with a small number of comments in the second headteacher questionnaire such as:

Workload excessive as before due to marking especially writing task... [marking] should be sent away as ks2

(headteacher)

and

Too much work. Try to reduce workload. My teachers were VERY conscientiousness about using both the test results and extensive teacher assessment.

(headteacher)

Both of these negative comments do not indicate unhappiness with teacher assessment being put to the fore *per se*, but rather, reflect disapproval of the implementation of the trial in their particular schools and the corresponding specific workload increases (or excessive workload not being reduced). This seems to fit a pattern of schools that appear educationally in agreement with the aims of the trial but find the practicalities very demanding, possibly due to not taking full advantage of the flexibility on offer and keeping with a high level of testing as in previous years (see similar comments in section 4.2).

LEA coordinators

LEA coordinators were asked in their second questionnaire about the benefits of the new arrangements and some typical responses indicating strong approval of the move towards teacher assessment were:

increased focus and awareness of the importance and use of teacher assessment

(LEA coordinator)

Seeing assessment as more than just 'test' performance

(LEA coordinator)

The LEA coordinators were unanimously in favour of roll-out:

Table 4.25 – LEA coordinator questionnaire (2) – support for roll-out

	Yes	No
12. Do you think that the trial arrangements should be made permanent from 2005? (n = 22)	22 (100)	0 (0)

though with one or two caveats:

But we NEED to know ASAP

(LEA coordinator)

This is with the proviso that the requirement for tests and tasks is increasingly seen to be more like an 'array of assessment tools' which teachers can use if and when they like and when it is most useful for children

(LEA coordinator)

Governors

There was a broad welcome in the governor questionnaire of the move towards teacher assessment with approximately 42% of governors stating that one benefit of the new arrangements would be that teacher assessment was more accurate and/or fairer [than previous arrangements]. Further, 30% of governors felt that the trial would raise the profile of teacher assessment and give teachers' judgements more status and credibility. A typically favourable observation was:

It should make the assessment procedure less stressful + more accurate representation of their abilities.

(governor)

There were one or two more double-edged remarks:

Whilst welcoming anything that de-emphasises SATs & the stresses they cause on children, families & the school, I await evidence that this will be an improvement.

(governor)

4.8 Key findings and issues

- The majority of stakeholders were positive about the trial, and there is strong support for roll-out.
- Views are, of course, varied both across and within LEAs, but there are no parts of the country where negative views are more than a small minority.
- All the respondents felt that the stress and pressure on the children had improved or stayed the same.
- The enhancement to teachers' professionalism was widely appreciated.
- There is a general agreement among the professionals that moderation is vital to the new system.
- Teacher confidence is critical, and improved as a result of taking part in the trial.
- Most professional respondents felt that despite a few minor difficulties and some obvious frustrations, the trial has been much more than a just a qualified success, and will improve further in future years.
- Parents seem by and large happy with the trial arrangements and there is strong evidence that they have faith in their own child's teachers' professional ability to assess their children accurately.
- There is evidence that as a consequence of the new arrangements there might be some increase in the amount of testing in Junior schools done early in Year 3.

5. Summary of findings and final discussion

Key stage 1 assessment was the first to be trialled and implemented after the introduction of the National Curriculum. This was well over a decade ago and now, as then, it arouses strong debate about the appropriateness of testing 7-year-olds, and the best ways of doing this. As such many of the issues raised in this evaluation are not new, although with the passage of time there are some changes of emphasis. Throughout the history of National Curriculum Assessment, at all key stages, there has always been a tension between the key issues of reliability, validity and manageability, and this is especially the case at key stage 1. The demands of assessment reliability rub up against those of validity in an uneasy way for this age group, and both are set against a background of ensuring that unreasonable demands should not be placed on teachers – i.e. manageability in the infant classroom.

The Introduction to this report emphasises the following points.

- KS1 assessments have varied historically, with some aspect changing almost every year. They are now much more pencil and paper orientated than when they were first introduced and hence, over time, a different balance has been struck between reliability, validity and manageability in these assessments.
- Recent investigations have shown that, on the whole, most teachers are happy with the quality and style of the tests/tasks, although the published research shows that some still think it is inappropriate formally to test 7-year-olds, and there may be some issues of the stress caused to young pupils.
- Over the years, many points have been raised about teacher assessment, and these are undoubtedly complex, especially in terms of the relationship, theoretical and practical, between this form of assessment and the national test results.
- Some schools do have a successful and supportive ‘assessment community’, although this does not usually extend across schools or LEA boundaries, which would need to be the case for Teacher Assessment in its new role to be reliable and robust.
- Evidence also suggests that, over time, test and Teacher Assessment levels have become very close at KS1.
- There are important questions to answer about the use of the various forms of assessment at key stage 1 and value-added calculations carried out at later points in progression through the key stages.

Against this background, the new proposals for Key Stage 1 assessments have emerged, and been trialled this year. This more flexible approach to testing, which uses national tests to underpin teacher assessment, represents a re-emphasis on the validity aspects of the assessments, provided that manageability factors are not compromised – i.e. workload.

The evaluation of this trial, reported here, was asked to address five questions:

- **Does the new system provide national data for value-added and target setting purposes that are as robust as the current system?**
- **What are the additional management and workload implications of such a system for teachers, schools and LEAs?**
- **What is the impact on, and views of parents, children and governors, as well as those of the school/LEA stakeholders of the new system compared to the existing one?**

Evaluation of the trial assessment arrangements for key stage 1

- **How valuable do schools and LEAs find the emphasis on Teacher Assessment in comparison to the previous system?**
- **Do parents find the new system easier to understand and does it give them enough information?**

In order to answer these questions, a sample of LEAs and schools was chosen from those taking part in the trial, and a wide range of appropriate data collected. This sample of LEAs and schools has been shown to be representative, and the numbers involved should ensure that the findings presented here have validity, depth and generalisability.

The findings on each of the evaluation questions are summarised in the following sections of this final summary and discussion.

5.1 Does the new system provide national data for value-added and target setting purposes that are as robust as the current system?

Chapter two emphasises that the notion of ‘robustness’ is a complex one. It is clear that any new system of assessment must be at least as robust as the present system in order to continue to support the monitoring of progress at both child and school level. The statutory tests and tasks at key stage 1 are made up of a complex set of assessments in English and mathematics, with strict rules about which tests and tasks are to be administered to which children. The tasks can be administered at any point over a long time period and the formal tests are not time-limited. The assessments are delivered and evaluated/marked by the teachers themselves. ‘Robustness’ in the present statutory system should be judged against the criteria of:

- high quality and well-trialled tests;
- effective training for both test and task delivery and marking;
- appropriate and sensitive preparation of the pupils;
- thorough audit of all processes and procedures.

Teacher Assessment is also part of the profile of national assessments at this key stage, delivered as a summative judgement at the end of Year 2, but based on experiences and data collected by teachers over a longer time period. ‘Robust’ Teacher Assessment in the present system, in the relevant subjects, will be based on teacher judgements soundly grounded in evidence, appropriately recorded and reported, and systematically moderated in line with national standards. By implication, in the case of the tests and tasks and Teacher Assessment, there is clearly scope for these criteria not to be met, although schools and LEAs work hard to ensure reliability and quality control.

Against this background, as outlined in Chapter 2, this evaluation has addressed the question of robustness in two ways: a comparative analysis of assessment data for 2003 and 2004, and an evaluation of the training and moderation provided for teachers and schools in 2004. The findings are summarised below.

- The data are generally comparable.
- All the differences that have been found, even where statistically significant, are small, and

Evaluation of the trial assessment arrangements for key stage 1

could be subsumed within normal cohort differences expected from year to year.

- There are more level 2s overall in all subjects (with fewer level 1s and/or 3s) than in either Teacher Assessments or test/task results in 2003. However, but there are also more 2s (and fewer 1s and/or 3s) in the test / task outcomes than there were in 2003, and the pattern of relative difference is maintained, i.e. higher mean levels for Teacher Assessments than for test/task results in Reading, and lower mean levels for Teacher Assessments than for test /task results in Writing and Mathematics.
- There is no direct evidence in the evaluation that gender differences have been affected by the new trial arrangements. Nor is there any evidence of an LEA effect. There is, however, some evidence that the trial arrangements have differential effects depending on age, with the differences between older and younger children smaller in Teacher Assessments in the trial arrangements than they are in test and task results.
- There was concern that teachers may rely too much on test/task scores in coming to Teacher Assessments. There is certainly a high correlation between the two scores, but given that they are two measures of the same underlying competence, this is to be hoped for, and if they were independent, would be evidence that both sources are robust. However, it might also indicate too much dependence.
- There is no evidence that schools used the discretionary aspect of Teacher Assessment in the trial arrangements to shape results in a way that would be helpful to their PANDA or value-added scores. The patterns of relationship between test/task results and Teacher Assessments are very similar in different types of school (Infant / First schools on the one hand and Primary schools on the other), and strongly reflect the patterns evident in the evaluation schools in the 'old' system.
- Confidence in coming to judgements using a full range of evidence (class work and test / task scores) is an issue for training and moderation. Many teachers reported that the moderation process helped their confidence in doing this.
- The trial appears to have been conducted with good quality training, effective moderation and conscientious teachers, and the performance data from the trial were comparable in most respects with that from the 'old' system as operated in the trial schools in 2003. As such this suggests that the new arrangements offer a system of assessment that is at least as robust as the system in which tasks and tests and Teacher Assessment are reported separately.

Having stated this overall conclusion, certain points need to be emphasised. The proposed new system places more reliance upon teachers and the quality of the assessments they make about the skills and capabilities of the children they teach on a daily basis. This may sustain or improve the validity of the assessments, but reliability can only be assured through rigorous training and moderation. Assessment theory and research indicate that, in systems where reliance is placed on a very large number of individual teachers making assessment judgements, the reliability of the assessments will be directly linked to the quality and effectiveness of the training and moderation that is put in place. This needs to ensure the same criteria and standards are being applied across all teachers, schools and (in the case of the system in England), across LEAs. It is not clear that such thorough and pervasive systems exist at present, especially from LEA to LEA. More will be said about training and moderation later in this section: only the major principle has been argued here.

But what of the issue of smaller numbers of children being assessed at Levels 1 and 3 in the trial? Chapter 2 has argued that fewer Level 1 assessments could be the result of the different context of the new assessments. Children on the Level 1/2C boundary may be more inclined to respond badly in a test situation and may under-perform in this one-off context, compared to their capabilities over time in class. The equivalent argument explaining the lower incidence of Level 3 outcomes is that, in the test situation, it is possible to garner enough marks from across a test to be awarded this level.

Evaluation of the trial assessment arrangements for key stage 1

But there are other arguments too. It could be suggested that teachers offer more help to children operating at Level 1, consciously or unconsciously, thus giving rise to this outcome. For the higher attaining children, the argument could be that the constraints of the curriculum and class work mean that more capable children may not be allowed to show their full Level 3 potential in their daily work.

The tensions between these arguments cannot be resolved by evidence and both are possibilities. The importance of tests and tasks as a standard national calibration should not mask the possibility that there are always uncertainties in individual cases. Equally, teachers' professional judgements may indeed provide valid insights into children's attainments, but there may be issues of bias, relativism and 'halo' effects (that is, generalising assessments from areas for which there is evidence of high achievement into areas for which there is not).

This points to the importance of checks and balances in the system. It could be that the proposed new system offers the best of both worlds, enabling teachers to act on doubts about specific test results, but with the test and task results themselves encouraging more balanced professional questioning of a range of assessment outcomes.

The evaluation has also shown that teacher confidence in making these judgements was generally hesitant at the beginning of the trial, since for many years there seems to have been a heavy reliance on test and task results in the overall outcomes. However, teachers' confidence grew throughout the trial as they understood and implemented the new arrangements. The expectation expressed by the teachers themselves, but also headteachers and LEA coordinators, is that as the arrangements become established, confidence will grow further, although a transition period may need to be allowed for.

Another key point in relation both to teacher confidence and the quality of the judgements made is the approach to assessment in each school. Where schools had well-established assessment policies and systems, then this period of adjustment was shorter, and confidence grew. These schools seemed to become comfortable with the new arrangements more rapidly. However, as other recent research has shown, not all schools have such well developed policies and practices, and hence the learning curve will be steeper for them, and the period of adjustment necessarily longer. The development of assessment policies and assessment procedures has not been high on the agenda of many primary schools, and this may need to be changed.

Training

The general significance of training has already been emphasised, but now more detail needs to be added. All schools need training in the assessment arrangements, whatever form they take, but in any period of change, the need for this is even greater. The findings of this evaluation suggest that, if roll-out happens in 2004/5, different kinds of training need to be offered to different groups. The training offered to 'lone' Year 2 teachers in a school may need to be different to that offered to groups of Year 2 teachers in larger schools. Teachers new to Year 2 may also need a different emphasis in their training to more experienced teachers. School managers may also need to be considered as a separate group. The focus for teachers, notwithstanding their different school circumstances and experience, needs to be on levelling and the quantity and quality of evidence: the focus for managers being on the wider system, the meaning of 'flexibility', and across school and LEA issues. Training also needs to be delivered early in the school year.

What has emerged from this evaluation is that the key message is clarity – clarity about the need for evidence and its quality and quantity; about just how 'flexible' the system is in terms of the range of choices open to the teacher; about criteria, levels and sub-levels; and about similarities and differences across subjects. There was some evidence in the trial that the notion of 'flexibility' was overplayed at the expense of conveying exactly what would be involved. In roll-out most schools

would be starting from scratch with the new system, and clear guidance needs to be given from the start.

Moderation

The importance of moderation has also been emphasised already in this summary, but once again the detail needs to be added. At present, many moderation systems exist in different LEAs, and this kind of range was reflected in both the trial LEAs and in the sub-set of these that took part in the evaluation. However, the evaluation has yielded few clear-cut examples of ‘good’ or ‘poor’ practice in moderation. By the end of the trial all the LEAs involved were identifying refinements and improvements for the following year, without suggesting that anything had been a mistake as such. The LEAs differ in size, school accessibility, socio-economic factors etc, and these circumstances must clearly be taken into account in offering advice should roll-out occur in 2004/5. Some flexibility of approach will be needed, but under the proposed new system, this will need to be much more limited, in the interests of both validity and reliability.

This said, some comments can be made about a few common themes within moderation practice that were positive. Moderation meetings need to be held early in the session in order to facilitate timely agreement on standards both within LEAs and across LEAs. This can be aided by the use of materials brought or sent in by schools for ‘levelling’. The moderation meetings seem to work best when the emphasis is on negotiation rather than imposition, whilst being clear about standards. Moderators appreciated meeting the children and not just looking at their work, although this needs to be handled with care and sensitivity. They need to be especially clear about the range and quantity of the evidence needed. It seemed to be important that LEA staff were available to follow up particular problems and ‘loose ends’, and to know when and how to offer additional support when it was needed. The strongest implication of the new arrangements, however, seems to be for schools themselves, supported by LEA staff, to develop their own assessment policy and understandings. It is the process of coming to this understanding that is important, and would be far preferable to the external imposition of standards that are poorly understood by those expected to implement them.

In any roll-out, there will need to be some clear thinking on the part of LEAs about how their staff can be helped (in many cases) to move from the ‘auditing’ role with schools to the ‘moderation’ role, that is from policing to support. In some cases this may represent a steep learning curve.

5.2 What are the additional workload and management implications of such a system for teachers, schools and LEAs?

The implications in terms of workload and management of the new arrangements arise partly from the specific changes that are involved, but also from the fact of change itself. The time and effort needed to learn new procedures, master them and embed them into new routines, are such that the simplification and reduction in work that the new system seems to promise (and which may be made explicit in training and information materials) does not manifest itself quickly. There is therefore a risk that the beneficial qualities of a change are not appreciated because of the need to adjust to it, and an unfair disillusion about the change by those who operationalise it overtakes the optimism of those who recommend it.

For that reason, there was a widespread concern at the outset about the new arrangements for assessment at key stage 1 that, however well meant or desirable they may be, they would be seen to add to the workload of teachers and schools, at least in the short term, and be difficult to manage.

Evaluation of the trial assessment arrangements for key stage 1

The conclusions of the evaluation of this aspect, reported in detail in Chapter 3, are broadly that the workload was not felt to be problematic, and that the management of the new arrangements was reasonably straightforward. Given the challenges that change of any kind brings, this is doubly reassuring, suggesting that not only are the specific changes involved in the new arrangements welcomed, they are welcomed to a sufficient extent to compensate for the inconvenience and inevitable costs of making adjustments to working practices.

At the heart of the reason for this is the observation that teachers felt the changes to be of benefit to the children in their care, and to enhance their own professionalism. There might well have been more time and effort involved, but this seems to have been discounted because it was felt to have been expended more clearly in relation to purposes that are part of their core teaching role. This may be contrasted with activities that relate to the purposes of others.

It should be noted, however, that the teachers taking part in the trial were all (at some level) volunteers, and in roll-out there may be schools that do not want the changes that the new arrangements represent, and so may resent whatever is involved in implementing them.

The findings of Chapter 3 on workload and management are summarised below.

- Assessment at key stage 1 presents a heavy burden for schools but the new arrangements have not been seen as adding to that burden significantly.
- Most teachers have reacted positively to the new arrangements. Many have recognised opportunities to reduce their workload.
- Others have been more cautious in the first run of the arrangements and admitted to doing more work than strictly necessary, though this may have brought necessary reassurance. These teachers see the prospect of workload reduction next time around.
- A minority have revealed some professional weakness that will need attention.
- LEA assessment coordinators have tended to see the new arrangements as an opportunity for teacher professional development. In this case any slight increase in workload arising from the greater personal involvement of moderation, compared to auditing tests and task outcomes, is seen as the price of good investment for the future.
- Data collection problems this year did increase the workload of LEA staff considerably, but this should be short term, and is in any case not directly related to the innovations of the new assessment arrangements.
- Management concerns before experience in implementing the new arrangements clustered around ensuring rigour and consistency, teacher confidence (with its corollary of over-preparation and excessive collection of evidence) and communication with parents.
- Schools with well-established assessment procedures seemed better placed to accommodate the new arrangements with minimum fuss.
- The management responsibility for ensuring robustness in the teachers' assessments may be more deeply felt in roll-out than it was in the trial.
- Teachers' confidence in their own assessment judgements was seen as vital, and although this will grow as a result of experience, some individual teachers and groups of teachers (e.g. those new to Year 2 in single entry schools), may need additional support from managers.
- Managing 'negative' feedback successfully (i.e. when the teachers' judgements are not supported by others or by test and task results) can be a critical point for teacher confidence.

Evaluation of the trial assessment arrangements for key stage 1

- Clear guidance about what evidence to keep in order to justify assessment judgements was seen as critical, whether for moderation purposes or generally, but some teachers still kept more than they were advised to do, especially at first, as a result of insecurity. This is expected to improve over time.
- School managers need to know early in the year about what is involved in order to organise the school resources and timetable appropriately.
- Greater flexibility does not mean total freedom, and school managers need clear guidance about where teachers are free to choose, and where they are not, in order to plan accordingly.

Some of these points have already been made and discussed in relation to the first research question, because they also have implications there. However, it is clear that there are issues and concerns to be faced here and lessons to be shared should there be roll-out in 2004/5.

As a final point, it became clear in the evaluation that whatever the perceived effects of changes to arrangements (and their attitudes to it) were, most teachers and other professionals agreed that the workload and management demands of KS1 assessment *per se* remain a large burden. Many, though not all, would seek to ease this by the complete removal of National Curriculum testing. A few would even dispute the need for any assessment at all at this age. If, however, the need for some form of assessment is accepted, then the subsequent demand for reliability and robustness would seem to imply a substantial amount of work, whatever the process.

5.3 The remaining research questions

- **What is the impact on, and views of parents, children and governors, as well as those of the school/LEA stakeholders of the new system compared to the existing one?**
- **How valuable do schools and LEAs find the emphasis on Teacher Assessment in comparison to the previous system?**
- **Do parents find the new system easier to understand and does it give them enough information?**

These three questions will be taken together, since to some extent the findings overlap. They are reported in Chapter 4 and are summarised here in the sections below.

5.3.1 Comparing the new arrangements to the old

5.3.2 The emphasis on teacher assessment

5.3.3 Parents' views

5.3.1 Comparing the new arrangements to the old

There was a general feeling across stakeholders that the new arrangements brought far more benefits than challenges in comparison to the old, and amongst heads, teachers and LEA staff there was overwhelming support for roll-out in 2004/5.

The most commonly stated benefits to the new arrangements included an expectation of an increased sense of professionalism amongst teachers and of a better classroom experience for children in Year 2, particularly a sense of less stress as formal testing at the end of key stage 1 moved into the background rather than being centre stage. Of course, such feelings were not universal and there was considerable evidence that the full flexibility of the trial had not (yet) been

Evaluation of the trial assessment arrangements for key stage 1

embraced in many schools and that as a consequence the experience of a significant number of Year 2 children, and possibly that of their teachers, had not changed very much. Overall, however, teachers were generally felt to be under reduced pressure as a consequence of the trial arrangements.

There was also a feeling, particularly from the LEA coordinators, that under the new arrangements the assessment regime was becoming more educationally relevant for the age group and that this dovetailed appropriately with other education issues that are currently to the fore (Assessment for Learning and so on). Finally, there was a strong feeling that the new arrangements did provide a more accurate measure of pupil performance, and that only reporting Teacher Assessments to parents was certainly a benefit.

In terms of the challenges that the arrangements brought, moderation was initially seen as key and, tied to this, the need for teacher judgements to be sufficiently justifiable to those interested in or affected by the results (parents, other teachers, schools, LEAs, government). However, as the trial proceeded, worries about moderation tended to recede and other issues appeared. These included the difficulty for teachers in getting used to using the full flexibility available to them under the new arrangements, and the confidence of teachers to be able to trust their own judgements. Data collection issues were also mentioned, but as noted above these were not specifically linked to the trial.

Often, many of the negative comments that were gathered during the course of the evaluation were not specifically about the trial arrangements, but were more generally about the formal testing of this age group. As testing remains an important part of assessment in the trial, the new arrangements were criticised by some for this, despite the fact that the tests results were no longer being reported and only used to inform Teacher Assessments.

5.3.2 The emphasis on teacher assessment

Throughout the course of the trial, it was difficult to obtain clear and definite information as to the effect that the trial arrangements generally, and the emphasis on Teacher Assessment in particular, would have on the use of the KS1 assessment outcomes. Junior schools revealed an increased inclination towards more internal testing in Year 3, though the new arrangements did not appear to be of great concern to them. The whole-school assessment implications of the trial were likewise difficult to ascertain and this is probably an area where further follow up research (under roll-out) is necessary as the new arrangements ‘bed-down’.

It may be that the new arrangements do not in themselves require schools to operate in a significantly different way, but they do seem to offer an opportunity for schools to develop a more integrated and holistic approach to assessment, in which formative and summative purposes are linked, and where the approach in Year 2 is more similar to that in Year 1 and Year 3 (and beyond). The LEA assessment coordinators who were interviewed all seemed of the view that this potential for development of whole-school assessment was present in the trial arrangements, and were keen to find ways to help schools exploit it – but were also aware of the risk that change can be approached conservatively, in terms of asking how little can be altered to meet the new requirements, rather than by asking in what ways the new arrangements can enable the school to move forward and enhance its provision.

5.3.3 Parents’ views

Most parents seemed quite happy with the trial arrangements. The majority view seemed to be that they had confidence and trust in their child’s teacher to assess their child’s progress

Evaluation of the trial assessment arrangements for key stage 1

accurately. The move away from formal tests at this age was generally welcomed, and the issue of stress and the possibility that fewer tests under the new arrangements would reduce this for their child was important to many parents.

However, there were also tensions within the views expressed. The common desire among parents to have both a good educational experience for their children was in a number of cases combined with a desire to have very full information from and about the school, in order to be reassured that the school is 'doing its job', and that the parent was correct in their choice of that school for their child. In the context of school assessment this often combined a desire for the children not to be tested with wanting to have complete test results available. There was no particular feeling that parents regretted the loss of the automatic reporting of test results for their own child, and only a small minority actually asked for these separately. However, they seemed to want to know the test results for the school, in order to make comparisons with other schools and the national picture. This is not possible, so there may be a perception issue that will need to be addressed.

This must also be understood in the context of what appears to be some confusion among parents about the assessment arrangements. The actual level of parental understanding of the KS1 assessment levels and what these really mean remains relatively poor, though teachers and heads did feel that the new arrangements were clearer and easier for parents to understand than the previous arrangements.

Another interesting tension within the views of parents was the combination of on the one hand an almost universal belief among them that their own child's teacher (and the school) is good, and (both) can be trusted to be fair and accurate in making assessments, with, on the other hand, a widespread feeling that teachers in other schools cannot be trusted to the same extent. In the context of school assessment, this often translated as the parent being comfortable about the teacher of their own child using his or her own judgement, but wanting other teachers to use an objective test (so that their own child's achievements – and those of the school - would not be undermined by unfair comparisons). A questionnaire comment along these lines was:

We are fortunate to have an excellent teacher who knows the children very well (we are a small school = ~ 150 pupils in primary). In larger schools with higher staff turnover I would be concerned that the teacher assessment would be unfair.

(parent)

Although not directly relevant to the arrangements being trialled, many parents expressed their hope of being told more about their child's progress than is provided by Teacher Assessment levels, with or without test / task results, at the end of the year. They want to be given information in order to help their child, and this implies being told of the results of assessment throughout the year, not just at its end. While this is not new, it could be that it would become more feasible if assessment within schools is developed along the lines that are envisaged for the new arrangements, with summative purposes linked to formative purposes, and where judgements about the levels at which children are working are integrated into the on-going assessment and record keeping systems of the school.

The main findings of Chapter 4 are summarised below.

- The majority of stakeholders were positive about the trial, and there is strong support for roll-out.
- Views are, of course, varied both across and within LEAs, but there are no parts of the country where negative views are more than a small minority.

Evaluation of the trial assessment arrangements for key stage 1

- The enhancement to teachers professionalism was widely appreciated.
- All the respondents felt that the stress and pressure on the children had diminished or stayed the same.
- There is a general agreement among the professionals that moderation is vital to the new system.
- Most professional respondents felt that despite a few initial problems and some obvious frustrations, the trial has been much more than a just a qualified success, and will improve further in future years.
- There is evidence that as a consequence of the new arrangements there might be some increase in the amount of testing in Junior schools done early in Year 3.
- Parents seem by and large happy with the trial arrangements and there is strong evidence that they have faith in their own child's teachers' professional ability to assess their children accurately.

References

- CFAS (2002) Analytical Summary to QCA on The evaluation of statutory assessment in 2002. Key Stage 1. The tests and administrative arrangements. Prepared by Centre for Formative Assessment Studies, Faculty of Education, University of Manchester.
- Changes to assessment 2003: Sample materials for key stages 1 and 2. QCA, London. www.qca.org.uk/ages3-14/.
- Chisholm, W. (2004) *The impact of high stakes national curriculum assessment on primary schools at key stage 2*. Unpublished Ph.D. thesis, University of Leeds.
- Connor, M J (2003) Pupils stress and standard assessment tasks (SATs): an update. *Emotional and Behavioural Difficulties*, 8, (2) May 2003, p.101-7.
- Durant, D. (2003) A comparative analysis of Key Stage tests and teacher assessments. Paper presented at the British Educational Research Association Annual Conference, Heriot-Watt University, Edinburgh, 11-13 September 2003.
- Gipps, C. (1995) Teacher assessment and teacher development. *Education*, 3, 1-8.
- Hall, K. and Harding, A (2002) Level descriptions and teacher assessment in England: towards a community of assessment practice. *Educational Research*, 44, 1, 1-15.
- Harlen, W., Gipps, C., Broadfoot, P., and Nutall, D. (1992) Assessment and the improvement of education. *Curriculum Journal*, 3, 215-230.
- NFER, Macaulay, A., Archer, T. & Kendall, L. (Spring 2003) Annual Survey of Trends in Education. Slough.
- Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174
- Rea-Dickins, P. (2001) Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing*, 18, 4, 429-462.
- Sainsburys, M and Sizmur, S. (1998) Level descriptions in the National Curriculum: what kind of criterion referencing is this?, *Oxford Review of Education*, 24, 2, 181-93.
- Sheskin, D. (2004) *Handbook of parametric and nonparametric statistical procedures* (3rd edition). Boca Raton. Chapman & Hall.
- Shorrocks, D., Daniels, S., Frobisher, L., Nelson, N., Waterson, A., and Bell, J. (1991) *ENCA 1 Project*. The Evaluation of National Curriculum Assessment at Key Stage 1.
- Shorrocks-Taylor, D. (1999) *National Testing: Past, Present and Future*. British Psychological Society. Leicester.
- Sizmur, S. and Burley, J. (1997) Assessment in the core subjects in 1996: Key Stage 1 Evaluation Report. London. School Curriculum and Assessment Authority.
- Streiner D.L. & Norman, G.R.(1995) *Health Measurement Scales* (2nd edition). New York. OUP.
- Task Group on Assessment and Testing (TGAT) (1988) A Report (London, Department of Education and Science).
- William, D. (1998) Enculturating learners into communities of practice: raising achievement through classroom assessment. Paper presented at the European Conference on Educational Research (ECER) Ljubljana, Slovenia, September.

Appendices

This Appendix includes copies of all the questionnaires as sent out to the various stakeholders followed by a comprehensive quantitative summary of all questionnaire responses. Also included are the summarised moderation plans and the set of questions making up the interview and focus groups schedules that were used in the evaluation.

The Appendix is divided into five parts:ERROR! NOT A VALID HEADING LEVEL RANGE.**Appendix A.** The questionnaires as administered

Appendix B. Quantitative questionnaire responses

Appendix C. Interview schedules

Appendix D. Focus group schedules

Appendix E. Moderation plan summaries

The questionnaires as administered

Table A.19 list details of the format of the questionnaires that were administered

Table A.19 - Details of the format of the questionnaires

Questionnaire	Format
Governor	2 page A4 back to back single sheet
Headteacher (1)	4 page A3 folded booklet
Headteacher (2)	4 page A3 folded booklet
Junior school	2 page A4 back to back single sheet
LEA coordinator (1)	4 page A3 folded booklet
LEA coordinator (2)	4 page A3 folded booklet
LEA other	2 page A4 back to back single sheet
Parent	2 page A4 back to back single sheet
Teacher (1)	4 page A3 folded booklet
Teacher (2)	4 page A3 folded booklet

The governor and parent questionnaires varied very slightly in that they were printed with the LEA name on them to avoid confusion. In what follows only Dorset's versions of these are shown as typical examples. The LEA was written in by the person completing the questionnaire in the other questionnaires, apart from, for reasons of anonymity, the LEA coordinator questionnaire.

The 'LEA other' questionnaire was administered in a variety of slightly modified forms to LEA personnel with responsibility for Early Years, Literacy, Numeracy, English and Maths. Only those personnel with direct responsibility for any of the national strategies were asked question 3 about the trial's impact on this area of their work. To avoid repetition only one example of this questionnaire, for a Literacy Line Manager, is shown in what follows.

The actual questionnaires are now presented in the order as shown in Table A.19.



**Key Stage 1 Assessment Trial Evaluation
Governor views in Dorset**

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions.

Your answers cannot be traced back to you. However, it would be helpful if you could tell us which type of governor you are, e.g. LEA, teacher, parent etc.

Type of governor: _____

Date: _____ 2004

Please read each question and give the answer that best fits your view.

1. Do you know about the Key Stage 1 assessment trial that is operating in your school this year? Yes No

If no, please read the information on the attached sheet before continuing.

2. Have the new arrangements been discussed in a Governors' meeting? Yes No

If yes, please give an indication of:

(a) the approximate date:

(b) what the issues were:

(c) what action was taken:

3. In the last few years, how often have issues about Key Stage 1 assessment arisen in Governors' meetings?

Frequently

Occasionally

Rarely

Never

Don't know

If any Key Stage 1 issues have been discussed, please say what they have been:

Please turn over to continue

4. In what way(s) do you think that the new approach to Key Stage 1 assessment is likely to bring benefits to the school?

5. In what way(s) do you think that the new approach to Key Stage 1 assessment is likely to bring new challenges for the school?

6. What impact do you expect the assessment trial will have on the workload of Y2 teachers?

7. To what extent do you expect the trial to affect the work of the head teacher / management team?

8. In relation to the trial what do you see as the greatest challenge to management?

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.

Please return this form in the envelope provided or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Thank you for your help



**Key Stage 1 Assessment Trial Evaluation
Headteacher views (1)**

LEA:

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions. We will contact you again in July 2004 with a follow-up questionnaire.

Your answers cannot be traced back to you. However, if you could tell us how many years you have been a teacher it would be very helpful.

Length of teaching experience:years

Please return the form in the envelope provided, or use this freepost address (no stamp is

required):

Assessment and Evaluation Unit
 11-14 Blenheim Terrace
 School of Education
 The University of Leeds
 FREEPOST LS 3018
 Leeds LS2 3YY

Please turn over to continue

Date: _____ 2004

Last Year (2002/3)

For the following questions, please give your opinion of the KS1 assessment arrangements employed last year (2002/3).

	Minimal	A small amount	A reasonable amount	A large amount	Too much
1. How much work was required in doing the assessments in 2002/3?	<input type="checkbox"/>				
2. How well-supported was the school last year in terms of information, training and guidance received from:	Very poorly	Poorly	Adequately	Well	Very well
(a) the LEA?	<input type="checkbox"/>				
(b) QCA (exemplar materials etc)?	<input type="checkbox"/>				

Rate how strongly you agree with the following statements concerning the 2002/3 arrangements:

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know
3. They provided consistent and comparable information on children's performance over the year.	<input type="checkbox"/>				

- | | | | | | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 4. They provided a good measure of children's attainment level at the end of the year. | <input type="checkbox"/> |
| 5. They had a positive effect on teaching and learning in the classroom. | <input type="checkbox"/> |
| 6. They gave meaningful information and feedback for parents on their children's progress. | <input type="checkbox"/> |
| | | Yes | No | Don't know | |
| 7. They were clear and well understood by teachers. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | | Yes | No | Don't know | |
| 8. They were clear and well understood by parents. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |

Other comments

Please make any further observations or give any other information about last year's KS1 assessment arrangements that you think is relevant to our evaluation.

This Year (2003/4)

For the following questions, please give your opinion of the KS1 assessment arrangements being trialled this year (2003/4).

9. In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring benefits to your school? *(Please tick all that apply)*

- (a) Increased professionalism among teachers
- (b) Increased clarity in outcomes
- (c) Increased confidence in outcomes
- (d) Improved communication with parents
- (e) Fewer parental questions/issues
- (f) Better experience for the children
- (g) Other *(please specify)*

Please write the letters of the two that you feel to be the most important

10. In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring new challenges for the school? *(Please tick all that apply)*

- (a) Greater need to formalise teacher assessments
- (b) Greater need for transparency
- (c) Greater need for moderation of assessment
- (d) Greater need to justify assessment procedures
- (e) Greater need to justify assessment judgements
- (f) Other *(please specify)*

Please write the letters of the two that you feel to be the most important

11. What impact do you expect the assessment trial will have on the workload of Y2 teachers?

- Large reduction
- Small reduction
- No change
- Small increase
- Large increase
- Don't know

12. To what extent do you expect the trial to affect the work of the head teacher / management team?

- Make it much harder
- Make it a little harder
- No change
- Make it a little easier
- Make it a lot easier
- Don't know

What do you see as the greatest challenge to management?

13. What do you see as the main issues in implementing the trial arrangements?

Other comments

Please make any further observations or give any other information about this year's KS1 assessment arrangements that you think is relevant to our evaluation.



**Key Stage 1 Assessment Trial Evaluation
Headteacher views (2)**

LEA:

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions. This is the follow-up questionnaire to the first questionnaire which we sent to you in April 2004.

Your answers cannot be traced back to you. However, if you could tell us how many years you have been a teacher it would be very helpful.

Length of teaching experience:.....years

It would also be useful if you could tell us if your school has been/is being moderated this year.

Yes/No (*please circle*)

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
 11-14 Blenheim Terrace
 School of Education
 The University of Leeds
 FREEPOST LS 3018
 Leeds LS2 3YY

Please turn over to answer questions

Date: _____ 2004

For the following questions, please give your opinion of the KS1 assessment arrangements that have been trialled this year (2003/4)

	Minimal	A small amount	A reasonable amount	A large amount	Too much	Don't know
1. How much work was required in doing the assessments?	<input type="checkbox"/>					

2. Did the school meet all the required deadlines? Yes No

How did this compare to last year? Easier About the same Harder

In relation to the assessments:

	A lot less	A little less	About the same	A little more	A lot more	Don't know
3. Compare the workload of the Y2 teachers this year with that of last year.	<input type="checkbox"/>					
4. Compare the workload experienced by you / your management team this year with that of last year.	<input type="checkbox"/>					

5. What School Management System do you use?

6. Who enters the assessment data onto the School Management System, e.g. school secretary?

How did their workload this year compare with last year?	A lot less	A little less	About the same	A little more	A lot more		Don't know
	<input type="checkbox"/>		<input type="checkbox"/>				

7. How well-supported was the school in terms of information, training and guidance from:	Very poorly	Poorly	Adequately	Well	Very well		Don't know
(a) the LEA?	<input type="checkbox"/>		<input type="checkbox"/>				
(b) QCA (exemplar materials etc)?	<input type="checkbox"/>		<input type="checkbox"/>				

If you have any suggestions as to how the support could be improved please give details here.

Please rate how strongly you agree overall with these statements concerning the trial assessment arrangements:

	Strongly disagree	Disagree	Agree	Strongly agree	Don't know
8. They provided consistent and comparable information on children's performance over the year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. They provided a good measure of children's attainment level at the end of the year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. They has had a positive effect on teaching and learning in the classroom.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. They gave meaningful information and feedback for parents on their children's progress.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Agree	Disagree	Don't know		
12. They were clear and well understood by parents.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. They were clear and well understood by teachers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. They were clear and well understood by children.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. For how many children was there a variation between test/task result and teacher assessment:	Few or none	Some	Many	All or almost all	Don't know
(a) in reading?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) in writing?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) in mathematics?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. How many parents specifically asked for their child's test/task results:	Few or none	Some	Many	All or almost all	Don't know
(a) in reading?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) in writing?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) in mathematics?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. How confident are you about the accuracy of the overall attainment levels produced:	Very confident	Quite confident	Not very confident	Not at all confident	Don't know
(a) in reading?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) in writing?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) in mathematics?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Do you think that the trial arrangements should be made permanent from 2005?					
<input type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> Don't know			

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.



**Key Stage 1 Assessment Trial Evaluation
Junior Schools**

LEA:

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions.

Your answers cannot be traced back to you.

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Date: _____ 2004

1. Please describe the use that was made in Year 3 of last year's Key Stage 1 assessment results sent to you by the Infant Schools.

2. Do you administer your own tests / assessments early in Year 3, rather than or as well as using the results from the Key Stage 1 assessments? *Yes* *No*

If yes, please explain why.

3. In the trial, and possibly from now on, the end of Key Stage 1 assessment will just be a Teacher Assessment.

Is this likely to affect your use of the assessment information? *Yes* *No*

If yes, please say in what way(s).

Specifically, are you more likely or less likely to do your own initial assessments?

Please explain why.

<i>More likely</i>	<i>Less likely</i>	<i>No different</i>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.

Continue overleaf if necessary

Thank you for
your help



**Key Stage 1 Assessment Trial Evaluation
LEA Coordinator (1)**

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions.

Your answers cannot be traced back to you.

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Date: _____ 2004

1. How do you expect the training and moderation workload for LEA personnel to compare with last year?

Considerably more A little more About the same A bit less Considerably less

2. How do you expect the data collection and analysis workload for LEA personnel to compare with last year?

Considerably more A little more About the same A bit less Considerably less

3. What impact do you expect the assessment trial will have on the workload of Y2 teachers?

Large reduction Small reduction No change Small increase Large increase Don't know

4. To what extent do you expect the trial to affect the work of the head teacher / management team?

Make it much harder Make it a little harder None Make it a little easier Make it a lot easier Don't know

What do you see as the greatest challenge to school management in the new assessment arrangements?

5. In which of these ways (if any) do you think that the new approach to Key Stage 1 assessment is likely to bring benefits to schools? (Please tick all that apply)

- | | |
|--|---|
| (a) Increased professionalism among teachers <input type="checkbox"/> | (b) Increased clarity in outcomes <input type="checkbox"/> |
| (c) Increased confidence in outcomes <input type="checkbox"/> | (d) Improved communication with parents <input type="checkbox"/> |
| (e) Fewer parental questions / issues <input type="checkbox"/> | (f) Better experience for the children <input type="checkbox"/> |

(g) Other (please specify)

Please write the letters of the two that you feel to be the most important.

6. In which of these ways (if any) do you think that the new approach to Key Stage 1 assessment is likely to bring new challenges for schools? (Please tick all that apply)

- | | |
|---|---|
| (a) Greater need to formalise teacher assessments <input type="checkbox"/> | (b) Greater need for transparency <input type="checkbox"/> |
| (c) Greater need for moderation of assessment <input type="checkbox"/> | (d) Greater need to justify assessment procedures <input type="checkbox"/> |
| (e) Greater need to justify assessment judgements <input type="checkbox"/> | |

(f) Other (please specify)

Please write the letters of the two that you feel to be the most important.

7. Are there any particular categories of schools that you think will experience particular difficulties in implementing the trial arrangements for Key Stage 1 assessment? ÷ Yes ÷ No

8. What impact do you expect the new arrangements will have on the number of occasions when the LEA overrules the school's assessment judgements?

Increase

Decrease

Stay the same

Don't know

Why?

9. Please outline your advice to schools about incorporating test and task results into teacher assessment.

10. What do you expect to be the main issues in your moderation of teachers' assessments?

11. Please indicate what advice (if any) you will offer to schools about reporting results to parents. For example, will you advise schools to report more than the statutory requirements?

12. Does the LEA intend to change its use of the outcomes from Key Stage 1 assessment in school analyses such as value-added comparisons? ⊃ Yes ⊃ No
If yes, please describe how.

13. What do you think are the advantages/risks and potential strengths/weaknesses of the new arrangements compared to the old (if any)?

Other comments

Please make any further observations or give other information that you think may be relevant to our evaluation

Continue overleaf if necessary

Thank you for your help



**Key Stage 1 Assessment Trial Evaluation
LEA Coordinator (2)**

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions. This is the follow-up questionnaire to the first questionnaire which we sent to you in April 2004.

Your answers cannot be traced back to you.

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Date: _____ 2004

1. **How did the training and moderation workload for LEA personnel compare with last year?**
 Considerably more A little more About the same A bit less Considerably less

2. How did the data collection and analysis workload for LEA personnel compare with last year?
 Considerably more A little more About the same A bit less Considerably less

3. What impact did the assessment trial have on the workload of Y2 teachers?
 Large reduction Small reduction No change Small increase Large increase Don't know

4. To what extent did the trial affect the work of the head teacher / management team?
 Made it much harder Made it a little harder No change Made it a little easier Made it a lot easier Don't know

What do you think was the greatest challenge to school management in the new assessment arrangements?

5. Are there any particular categories of schools that experienced particular difficulties in implementing the trial arrangements for Key Stage 1 assessment? ÷ Yes ÷ No
If yes, please specify.

6. How confident were teachers about the accuracy of the overall attainment levels produced?
 Very confident Quite confident Not very confident Not at all confident Don't know

7. How confident were you about the accuracy of the overall attainment levels produced by teachers?

8. What impact did the new arrangements have on the number of occasions when the LEA overruled the school's assessment judgements?
 Increased Decreased Stayed the same Don't know

Why was this?

9. What were the main issues in your moderation of teachers' assessments?

10. What do you think have been the main benefits to schools of the new assessment arrangements?

11. What aspects of the new arrangements have schools found most challenging?

12. Do you think that the trial arrangements should be made permanent from 2005? Yes No Don't know

Other comments

Please make any further observations or give other information that you think may be relevant to our evaluation

Continue overleaf if necessary

Thank you for your help



**Key Stage 1 Assessment Trial Evaluation
Literacy Line Manager**

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions.

Your answers cannot be traced back to you.

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Please turn over to answer questions

1. What are the issues for the NLS in the Key Stage 1 assessment trial?

2. Will you be involved in the training / moderation of KS1 assessment? Yes

No

If yes, please give details.

3. Is there anything in the new system for assessment at Key Stage 1 that conflicts with the advice about assessment given in the NLS?

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.

Thank you for your help



**Key Stage 1 Assessment Trial Evaluation
Views of parents of children in Year 2 in Dorset**

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of schools in England. Your child's school is one of these schools.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions.

Your answers cannot be traced back to you.

Please read each question and give the answer that best fits what you think.

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Date: _____ 2004

1. Have you received the Teacher Assessment (TA) levels for your child this year? Yes No Don't know

2. Have you had enough information about the Teacher Assessment levels? Yes No Don't know

3. Where did you get most of the information about the Teacher Assessments from?

Speaking to the teacher Speaking to other Letters from the teacher Government

Other(s) (please say where
from)

4. Were the Teacher Assessment levels easy to understand? Yes No Don't know

5. Did you agree with all of the Teacher Assessment levels that your child got? Yes No

If no, please tick the ones you agreed with.

Speaking & Reading Writing Maths Science I agreed with none

6. Did your child agree with the Teacher Assessment levels that he/she got? Yes No Don't know

If no, please give more details.

7. Was your child worried about receiving the Teacher Assessment levels? Yes No Don't know

If yes, please give more details.

8. Did you ask for any of your child's Test levels? Yes No

If No ➡ Why not? (please tick all boxes that apply)

The Teacher Assessment I don't think the teacher asked I didn't know I I didn't know

Other(s) (please say why)

If Yes ➡ Which ones? (please tick all boxes that apply)

Reading Writing Spelling Mathematics

Why did you ask for these Test Levels? (please tick all boxes that apply)

I like to have all I wasn't satisfied with the I didn't think the Teacher Assessment



Other(s) (please say why)

9. Do you have any other children who have been assessed at Key Stage 1 in previous years? Yes No Don't know

If Yes, do you think that this year's arrangements are better? Yes No Don't know

*If you want to add anything else about the way Key Stage 1 is assessed, please use the space below.
(Please do not name any people)*

Continue overleaf if necessary
your help

Thank you for



**Key Stage 1 Assessment Trial Evaluation
Teacher views (1)**

LEA:

Changes to the system of national assessment of 7 year-olds are being trialled in about a quarter of LEAs in England. Your LEA is one of them.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions. We will contact you again in July 2004 with a follow-up questionnaire.

Your answers cannot be traced back to you. However, if you could tell us how many years you have been a primary teacher it would be very helpful.

Length of primary teaching experience:.....years.

Please return the form in the envelope provided, or use this freepost address (no stamp

is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Please turn over to answer questions

Date: _____ 2004

Last Year (2002/3)

For the following questions, please give your opinion of the KS1 assessment arrangements employed last year (2002/3).

(If you were not involved in KS1 assessments last year, please ask an appropriate colleague to complete this part. You can then complete the second page.)

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | Minimal | Small amount | Reasonable amount | Large amount | Too much |
| 1. How much work was required in doing the assessments in 2002/3? | <input type="checkbox"/> |
| 2. How well-supported were you in terms of information, training and guidance that you received from: | Very poorly | Poorly | Adequately | Well | Very well |
| (a) your school? | <input type="checkbox"/> |
| (b) your LEA? | <input type="checkbox"/> |
| (c) QCA (exemplar materials etc)? | <input type="checkbox"/> |

Give your opinion of the following statements concerning the 2002/3 assessment arrangements:

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | Strongly disagree | Disagree | Agree | Strongly agree | Don't know |
| 3. They provided consistent and comparable information on children's performance over the year. | <input type="checkbox"/> |
| 4. They provided a good measure of children's attainment levels at the end of the year. | <input type="checkbox"/> |
| 5. They had a positive effect on teaching and learning in the classroom. | <input type="checkbox"/> |
| 6. They gave meaningful information and feedback for parents on their children's progress. | <input type="checkbox"/> |
| 7. They were clear and well understood: | Yes | No | Don't know | | |
| (a) by me. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | |
| (b) by parents. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | |
| (c) by children. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | |

8. For how many children did you use test/task results to help you to arrive at TA levels?

Subject	All children	No children	Some children <i>(please specify which kinds of</i>
Reading	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mathematics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other comments

Please make any observations or give other information about last year's Key Stage 1 assessment arrangements that you think is relevant to our evaluation.

This Year (2003/4)

For the following questions, please give your opinion of the KS1 assessment arrangements being trialled this year (2003/4) in your school.

- | | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | A lot less | A little less | About the same | A little more | A lot more | Don't know |
| 9. In doing the assessments, compare the workload you expect this year against that required last year. | <input type="checkbox"/> |

Rate how strongly you agree overall with the following statements concerning the trial arrangements:

- | | | | | | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | Strongly disagree | Disagree | Agree | Strongly agree | Don't know |
| 10. They will provide consistent and comparable information on children's performance over the year. | <input type="checkbox"/> |
| 11. They will provide a good measure of children's attainment level at the end of the year. | <input type="checkbox"/> |
| 12. They will have a positive effect on teaching and learning in the classroom. | <input type="checkbox"/> |
| 13. They will give meaningful information and feedback for parents on their children's progress. | <input type="checkbox"/> |

- | | | | |
|---|--------------------------|--|--|
| 14. To what extent are you keeping examples of children's work this year: | Not keeping examples | Keeping one piece per week in each subject | Keeping several examples of work for each child in each area |
| (a) <i>writing?</i> in | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Other arrangement (please specify) | <hr/> | | |
| (b) <i>mathematics?</i> in | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Other arrangement (please specify) | <hr/> | | |

- | | | | | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|
| | Yes, all | Yes, | No, none | Don't know |
| 15. Have you attended (or will you be attending) your LEA's assessment/moderation training sessions for Year 2 teachers? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

16. How (if at all) are you using or do you intend to use the QCA trial exemplification materials which were sent to LEAs for them to use in training sessions?

17. Overall, I am clear about what is involved in the new assessment arrangements. Yes No

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.

Continue overleaf if necessary
your help

Thank you for



**Key Stage 1 Assessment Trial Evaluation
Teacher views (2)**

LEA:

The proposed new Key Stage 1 assessment arrangements are being trialled in your LEA.

We are carrying out the evaluation of the trial.

We would very much appreciate you giving us your views by answering some questions. This is the follow-up questionnaire to the first questionnaire which we sent to you in April 2004.

Your answers cannot be traced back to you. However, if you could tell us how many years you have been a primary teacher it would be very helpful.

Length of primary teaching experience:.....years.

It would also be useful if you could tell us if your school has been/is being moderated this year.

Yes/No (please circle)

Please return the form in the envelope provided, or use this freepost address (no stamp is required):

Assessment and Evaluation Unit
11-14 Blenheim Terrace
School of Education
The University of Leeds
FREEPOST LS 3018
Leeds LS2 3YY

Please turn over to answer questions

Date: _____ 2004

For all of the following questions, please give your opinion of the KS1 assessment arrangements that have been trialled this year (2003/4) in your school.

	Minimal	Small amount	Reasonable amount	Large amount	Too much	Don't know
1. How much work was required in doing the assessments?	<input type="checkbox"/>					

2. Did you meet all your required deadlines? Yes No

How did this compare to last year? Easier About the same Harder

	A lot less	A little less	About the same	A little more	A lot more	Don't know
3. Compare the workload you experienced this year against that last year.	<input type="checkbox"/>					
4. Compare the workload experienced by teaching assistants this year against that last year.	<input type="checkbox"/>					

What did you ask the teaching assistants to do?

5. To what extent did you keep examples of children's work?	Didn't keep examples	Kept one piece per week in each subject	Kept several examples of work for each child in each area
---	----------------------	---	---

(a) In writing

Other arrangement (please specify)

(b) In mathematics

Other arrangement (please specify)

6. How (if at all) did you use the QCA exemplification materials?

7. How well-supported were you in terms of information, training and guidance that you received from:	Very poorly	Poorly	Adequately	Well	Very well	Don't know
(a) your school?	<input type="checkbox"/>					
(b) your LEA?	<input type="checkbox"/>					
(c) QCA (exemplar materials etc)?	<input type="checkbox"/>					

8. Please tick which tests/tasks you used and tick or write in which years they were from.

<i>Reading</i>	<input type="checkbox"/> Level 1 task	<input type="checkbox"/> Level 2 task	<input type="checkbox"/> Level 2 test	<input type="checkbox"/> Level 3 test
	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004			
<i>Writing</i>	<input type="checkbox"/> Short task	<input type="checkbox"/> Long task	<input type="checkbox"/> Spelling test	
	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004	
<i>Mathematics</i>	<input type="checkbox"/> Level 1 task	<input type="checkbox"/> Level 2 test	<input type="checkbox"/> Level 3 test	
	year	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004	<input type="checkbox"/> 2003 <input type="checkbox"/> 2004	

Give your opinion of the following statements concerning the trial assessment arrangements:

		Strongly disagree	Disagree	Agree	Strongly agree	Don't know
9.	They provided consistent and comparable information on children's performance over the year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	They provided a good measure of children's attainment level at the end of the year.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11.	They had a positive effect on teaching and learning in the classroom.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12.	They gave meaningful information and feedback for parents on their children's progress.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13.	They were clear and well understood:	Agree	Disagree			Don't know
	(d) by me.	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
	(e) by parents.	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
	(f) by children.	<input type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>
14.	For how many children was there a variation between the level achieved in the test/task and the final reported level (teacher assessment):	None or almost none		Some	Many	All or almost all
	(a)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(b)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(c)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15.	How many parents specifically asked for their child's tests results:	None or almost none		Some	Many	All or almost all
	(a)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(b)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(c)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16.	How confident were you in the judgements you had to make when balancing tests, tasks and children's work to produce one overall attainment level for each child:	Not at all confident	Not very confident	Quite confident	Very confident	Don't know
	(a)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(b)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	(c)	i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17.	Do you think that the trial arrangements should be made permanent from 2005?					
	<input type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> Don't know			

Other comments

Please make any further observations or give any other information that you think may be relevant to our evaluation.

Continue overleaf if necessary

Thank you for your help

Quantitative questionnaire responses

This section contains a full account of the quantitative responses to the questionnaires. Missing responses ('No answer' and 'Don't know') are only tabulated if non-zero over a complete column. The results are presented in the order as shown in Table A.19 on page A111.

Governor questionnaire

162 questionnaires analysed

Table B.1- Governor questionnaire - Q1 – Awareness of trial arrangements

Do you know about the KS1 assessment trial that is operating in your school this year?	Yes	No	No answer
	151 (93)	9 (6)	2 (1)

Table B.20 - Governor questionnaire - Q2 – Discussion of trial arrangements

Have the new arrangements been discussed in a Governors' meeting?	Yes	No	No answer
	124 (77)	30 (19)	8 (5)

Table B.21 - Governor questionnaire – Q3 – Discussion of KS1

In the last few year, how often have issues about Key Stage 1 assessment arisen in Governors' meetings?	Frequently	Occasionally	Rarely	Never	Don't know	No answer
	64 (40)	53 (33)	19 (12)	2 (1)	19 (12)	5 (3)

Headteacher (1) questionnaire

83 questionnaires analysed.

Note that the first part of this questionnaire (questions 1 to 8) concern the KS1 assessment arrangements in 2002/3, not the trial arrangements of 2003/4.

Table B.22 - Headteacher (1) questionnaire - Q1 – Amount of work in the pre-trial arrangements

How much work was required in doing the assessments (in 2002/3)?	Minimal	A small amount	A reasonable amount	A large amount	Too much	No answer
	0 (0)	2 (2)	18 (22)	50 (60)	10 (12)	3 (4)

Table B.23 - Headteacher (1) questionnaire – Q2 – Levels of pre-trial support

How well-supported (in 2002/3) was the school last year in terms of information, training and guidance received from:	Very poorly	Poorly	Adequately	Well	Very well	No answer
	0 (0)	0 (0)	25 (30)	37 (45)	17 (20)	4 (5)
(a) the LEA?	0 (0)	0 (0)	25 (30)	37 (45)	17 (20)	4 (5)
(b) QCA (exemplar materials etc)?	0 (0)	4 (5)	32 (39)	34 (41)	6 (7)	7 (8)

Table B.24 - Headteacher (1) questionnaire – Q3-6 – Views on the pre-trial arrangements

Rate how strongly you agree with the following statements concerning the 2002/3 arrangements:	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer
3. They provided consistent and comparable information on children's performance over the year.	8 (10)	28 (34)	43 (52)	2 (2)	0 (0)	2 (3)
4. They provided a good measure of children's attainment levels at the end of the year.	2 (2)	28 (34)	46 (55)	4 (5)	0 (0)	3 (4)
5. They had a positive effect on teaching and learning in the classroom.	22 (27)	40 (48)	16 (19)	1 (1)	1 (1)	3 (4)
6. They gave meaningful information and feedback for parents on their children's progress.	12 (15)	42 (51)	25 (30)	2 (2)	0 (0)	2 (3)

Table B.25 - Headteacher (1) questionnaire – Q7, 8 – Understanding of the pre-trial arrangements

The 2002/3 arrangements were:	Yes	No	Don't know	No answer
7. clear and well understood by teachers.	76 (92)	4 (5)	1 (1)	2 (2)
8. clear and well understood by parents.	24 (29)	43 (52)	13 (16)	3 (4)

Table B.26 - Headteacher (1) questionnaire – Q9 – Benefits of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring benefits to your school?	Yes	No answer	Most important ⁹
Increased professionalism among teachers	58 (70)	23(28)	34
Increased clarity in outcomes	43 (52)	38 (46)	13
Increased confidence in outcomes	57 (69)	24 (29)	26
Improved communication with parents	22 (27)	59 (71)	0
Fewer parental questions/issues	18 (22)	63 (76)	4
Better experience for the children	61 (73)	20 (24)	42

⁹© Qualifications and Curriculum Authority 2004

Table B.27 - Headteacher (1) questionnaire – Q10 – Challenges of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring new challenges for the school?	Yes	No answer	Most important ⁹
Greater need to formalise teacher assessments	37 (45)	45 (54)	15
Greater need for transparency	24 (29)	58 (70)	9
Greater need for moderation of assessment	65 (78)	17 (21)	49
Greater need to justify assessment procedures	32 (39)	50 (60)	9
Greater need to justify assessment judgements	48 (58)	34 (41)	29

Table B.28 - Headteacher (1) questionnaire – Q11 – Impact on teacher workload

What impact do you expect the assessment trial will have on the workload of Y2 teachers?	Large reduction	Small reduction	No change	Small increase	Large increase	Don't know	No answer
	4 (5)	25 (30)	24 (29)	19 (23)	7 (8)	2 (2)	2 (2)

Table B.29 - Headteacher (1) questionnaire – Q12 – Impact on management workload

To what extent do you expect the trial to affect the work of the head teacher/management team?	Make it much harder	Make it a little harder	No change	Make it a little easier	Make it a lot easier	Don't know	No answer
	2 (2)	19 (23)	51 (61)	8 (10)	0 (0)	2 (2)	1 (1)

Headteacher (2) questionnaire

63 questionnaires analysed

Table B.30 - Headteacher (2) questionnaire - Q1 – Amount of work in the trial arrangements

How much work was required in doing the assessments?	Minimal	A small amount	A reasonable amount	A large amount	Too much
	0 (0)	1 (2)	31 (49)	29 (46)	2 (3)

Table B.31 - Headteacher (2) questionnaire – Q2 – Meeting deadlines

Did the school meet all the required deadlines?	Yes	No
	62 (98)	1 (2)

How did this compare to last year?	Easier	About the same	Harder	No answer
	9 (14)	42 (67)	9 (14)	3 (5)

Table B.32 - Headteacher (2) questionnaire – Q3 – Comparing teacher workload

In relation to the assessments, compare the workload of the Y2 teachers this year with that of last year.	A lot less	A little less	About the same	A little more	A lot more	No answer
	1 (2)	11 (17)	34 (54)	10 (16)	6 (10)	1 (2)

Table B.33 - Headteacher (2) questionnaire – Q4 – Comparing management workload

In relation to the assessments, compare the workload experienced by you / your management team this year with that of last year.	A lot less	A little less	About the same	A little more	A lot more	No answer
	2 (3)	4 (6)	44 (70)	12 (19)	0 (0)	1 (2)

Table B.34 - Headteacher (2) questionnaire – Q6 – Comparing administration workload

How did their [the person entering assessment data] workload this year compare with last year?	A lot less	A little less	About the same	A little more	A lot more	No answer
	1 (2)	3 (5)	36 (57)	14 (22)	7 (11)	2 (3)

Table B.35 - Headteacher (2) questionnaire – Q7 – External levels of trial support

How well-supported was the school in terms of information, training and guidance from:	Very poorly	Poorly	Adequately	Well	Very well	Don't know	No answer
	(a) the LEA?	1 (2)	4 (6)	30 (48)	18 (29)	10 (16)	0 (0)
(b) QCA (exemplar materials etc)?	0 (0)	3 (5)	36 (57)	14 (22)	5 (8)	3 (5)	2 (3)

Table B.36 - Headteacher (2) questionnaire – Q8-11 – Views on the trial arrangements

Please rate how strongly you agree overall with these statements concerning the trial assessment arrangements:	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer
	8. They provided consistent and comparable information on children's performance over the year.	0 (0)	2 (3)	46 (73)	14 (22)	0 (0)
9. They provided a good measure of children's attainment levels at the end of the year.	0 (0)	3 (5)	43 (68)	17 (27)	0 (0)	0 (0)
10. They had a positive effect on teaching and learning in the classroom.	1 (3)	12 (19)	31 (49)	10 (16)	6 (10)	3 (5)
11. They gave meaningful information and feedback for parents on their children's progress.	1 (2)	7 (11)	42 (67)	10 (16)	2 (3)	1 (2)

Table B.37 - Headteacher (2) questionnaire – Q12-14 – Understanding of the trial arrangements

The trial arrangements were:	Agree	Disagree	Don't know	No answer
12. clear and well understood by parents.	23 (37)	12 (19)	25 (40)	3 (5)
13. clear and well understood by teachers.	58 (92)	3 (5)	1 (2)	1 (2)
14. clear and well understood by children.	42 (67)	9 (14)	10 (16)	2 (3)

Table B.38 - Headteacher (2) questionnaire – Q15 – Variations between test/task and TA

For how many children was there a variation between test/task result and teacher assessment:	Few or none	Some	Many	All or almost all	Don't know	No answer
(a) in reading?	42 (67)	15 (24)	2 (3)	2 (3)	1 (2)	1 (2)
(b) in writing?	42 (67)	17 (27)	1 (2)	2 (3)	1(2)	0 (0)
(c) in mathematics?	44 (70)	15 (24)	1 (2)	2 (3)	1 (2)	0 (0)

Table B.39 - Headteacher (2) questionnaire – Q16 – Parental demand for test/task results

How many parents specifically asked for their child's test/task results:	Few or none	Some	Many	All or almost all	Don't know	No answer
(a) in reading?	47 (75)	4 (6)	3 (5)	3 (5)	2 (3)	4 (6)
(b) in writing?	47 (75)	5 (8)	2 (3)	3 (5)	2 (3)	4 (6)
(c) in mathematics?	47 (75)	5 (8)	2 (3)	3 (5)	2 (3)	4 (6)

Table B.40 - Headteacher (2) questionnaire – Q17 – Confidence in results

How confident are you about the accuracy of the overall attainment levels produced:	Very confident	Quite confident	Not very confident	Not at all confident
(a) in reading?	41 (65)	22 (35)	0 (0)	0 (0)
(b) in writing?	37 (59)	24 (38)	2 (3)	0 (0)
(c) in mathematics?	39 (62)	21 (33)	3 (5)	0 (0)

Table B.41 - Headteacher (2) questionnaire – Q18 – Views on roll-out

Do you think that the trial arrangements should be made permanent from 2005?	Yes	No	Don't know
	49 (78)	5 (8)	9 (14)

Junior school questionnaire

27 questionnaires analysed

Table B.42 – Junior school questionnaire – Q2 – Administration of own tests early in Year 3

Do you administer your own tests/assessments early in Year 3, rather than or as well as using the results from the Key Stage 1 assessments?	Yes	No
	21 (78)	6 (22)

Table B.43 – Junior school questionnaire – Q3 – Effect of trial on use of KS1 results

In the trial, and possibly from now on, the end of Key Stage 1 assessments will just be a Teacher Assessment. Is this likely to affect your use of the assessment information?	Yes	No	No answer
	11 (41)	15 (56)	1 (4)
Specifically, are you more likely or less likely to do your own initial assessments?	More likely	Less likely	No different
	9 (33)	1 (4)	17 (63)

LEA coordinator (1) questionnaire

26 questionnaires analysed

Table B.44 – LEA coordinator (1) questionnaire – Q1 – Training and moderation workload

How do you expect the training and moderation workload for LEA personnel to compare with last year?	Considerably more	A little more	About the same	A bit less	Considerably less
	5 (19)	10 (39)	9 (35)	1 (4)	1 (4)

Table B.45 – LEA coordinator (1) questionnaire – Q2 – Data collection and analysis workload

How do you expect the data collection and analysis workload for LEA personnel to compare with last year?	Considerably more	A little more	About the same	A bit less	Considerably less
	4 (15)	4 (15)	14 (54)	3 (12)	1 (4)

Table B.46 – LEA coordinator (1) questionnaire – Q3 – Year 2 teacher workload

What impact do you expect the assessment trial will have on the workload of Y2 teachers?	Large reduction	Small reduction	No change	Small increase	Large increase
	1 (4)	11 (42)	9 (35)	4 (15)	1 (4)

Table B.47 – LEA coordinator (1) questionnaire – Q4 – Headteacher/management workload

To what extent do you expect the trial to affect the work of the head teacher / management team?	Make it much harder	Make it a little harder	None	Make it a little easier	Make it a lot easier	Don't know
	0 (0)	5 (19)	13 (50)	6 (23)	1 (4)	1 (4)

Table B.48 - LEA coordinator (1) questionnaire – Q5 – Benefits of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring benefits to schools?	Yes	No answer	Most important ¹⁰
Increased professionalism among teachers	26 (100)	0 (0)	20
Increased clarity in outcomes	16 (62)	10 (39)	3
Increased confidence in outcomes	18 (69)	8 (31)	0
Improved communication with parents	5 (19)	21 (81)	1
Fewer parental questions/issues	7 (27)	19 (73)	0
Better experience for the children	26 (100)	0 (0)	18

Table B.49 - LEA coordinator (1) questionnaire – Q6 – Challenges of the trial arrangements

In which of these ways (if any) do you think that the proposed KS1 assessment arrangements are likely to bring new challenges for schools?	Yes	No answer	Most important ¹⁰
Greater need to formalise teacher assessments	16 (62)	10 (39)	8
Greater need for transparency	10 (39)	16 (62)	5
Greater need for moderation of assessment	25 (96)	1 (4)	17
Greater need to justify assessment procedures	13 (50)	13 (50)	3
Greater need to justify assessment judgements	15 (58)	11 (42)	4

Table B.50 – LEA coordinator (1) questionnaire – Q7 – Categories of schools hindered by the trial arrangements

Are there any particular categories of schools that you think will experience particular difficulties in implementing the trial arrangements for Key Stage 1 assessment?	Yes	No	No answer
	6 (23)	19 (73)	1 (4)

Table B.51 – LEA coordinator (1) questionnaire – Q8 – Impact on the overruling of school judgements

What impact do you expect the new arrangements will have on the number of occasions when the LEA overrules the school’s assessment judgement?	Increase	Decrease	Stay the same	Don’t know
	0 (0)	3 (12)	18 (69)	5 (19)

Table B.52 – LEA coordinator (1) questionnaire – Q12 – Impact on the use of outcomes

Does the LEA intend to change its use of the outcomes from Key Stage 1 assessment in school analyses such as value-added comparisons?	Yes	No	No answer
	4 (15)	15 (58)	7 (27)

LEA coordinator (2) questionnaire

22 questionnaires analysed

¹⁰© Qualifications and Curriculum Authority 2004

Table B.53 – LEA coordinator (2) questionnaire – Q1 – Training and moderation workload

How did the training and moderation workload for LEA personnel compare with last year?	Considerably more	A little more	About the same	A bit less	Considerably less
	3 (14)	14 (64)	2 (9)	3 (14)	0 (0)

Table B.54 – LEA coordinator (2) questionnaire – Q2 – Data collection and analysis workload

How did the data collection and analysis workload for LEA personnel compare with last year?	Considerably more	A little more	About the same	A bit less	Considerably less	No answer
	13 (59)	4 (18)	1 (5)	2 (9)	1 (5)	1 (5)

Table B.55 – LEA coordinator (2) questionnaire – Q3 – Year 2 teacher workload

What impact did the assessment trial have on the workload of Y2 teachers?	Large reduction	Small reduction	No change	Small increase	Large increase	Don't know	No answer
	0 (0)	9 (41)	5 (23)	4 (18)	0 (0)	1 (5)	3 (14)

Table B.56 – LEA coordinator (2) questionnaire – Q4 – Headteacher/management workload

To what extent did the trial affect the work of the head teacher / management team?	Made it much harder	Made it a little harder	No change	Made it a little easier	Made it a lot easier	Don't know	No answer
	0 (0)	1 (5)	14 (64)	3 (14)	0 (0)	3 (14)	1 (5)

Table B.57 – LEA coordinator (2) questionnaire – Q5 – Categories of schools hindered by the trial arrangements

Are there any particular categories of schools that experienced particular difficulties in implementing the trial arrangements for Key Stage 1 assessment?	Yes	No
	7	15

Table B.58 - LEA coordinator (2) questionnaire – Q6 – Teacher confidence in the results

How confident were teachers about the accuracy of the overall attainment levels produced?	Very confident	Quite confident	Not very confident	Not at all confident
	6 (27)	16 (73)	0 (0)	0 (0)

Table B.59 - LEA coordinator (2) questionnaire – Q7 – Their confidence in the results

How confident were you about the accuracy of the overall attainment levels produced by teachers?	Very confident	Quite confident	Not very confident	Not at all confident
	8 (36)	14 (64)	0 (0)	0 (0)

Table B.60 – LEA coordinator (2) questionnaire – Q8 – Impact on overruling of school judgements

What impact did the new arrangements have on the number of occasions when the LEA overruled the school's assessment judgements?	Increased	Decrease	Stay the same
	1 (5)	9 (41)	12 (55)

Table B.61 - LEA coordinator (2) questionnaire – Q12 – Views on roll-out

Do you think that the trial arrangements should	Yes	No
	22 (100)	0 (0)

LEA other questionnaire

22 questionnaires analysed. This questionnaire was essentially qualitative in nature and so no results are presented here.

Parent questionnaire

570 questionnaires analysed.

Table B.62 - Parent questionnaire – Q1 – Receiving of assessments

Have you received the Teacher Assessment (TA) levels for your child this year?	Yes	No	Don't know	No answer
	536 (94)	20 (4)	13 (2)	1 (0)

Table B.63 - Parent questionnaire – Q2 – Teacher assessment information

Have you had enough information about the Teacher Assessment levels?	Yes	No	Don't know	No answer
	393 (69)	130 (23)	40 (7)	6 (1)

Table B.64 - Parent questionnaire – Q3 – Information sources for assessments

Where did you get most of the information about the Teacher Assessments from? ¹¹	Speaking to the class teacher	Speaking to other parents	Letters from the school	Government publications	Other
	245 (43)	53 (9)	293 (51)	16 (3)	67 (11)

Table B.65 - Parent questionnaire – Q4 – Parental understanding of the assessments

Were the Teacher Assessment levels easy to understand?	Yes	No	Don't know	No answer
	458 (80)	70 (12)	28 (5)	14 (3)

Table B.66 - Parent questionnaire – Q5 – Parental agreement with the assessments

Did you agree with all of the Teacher Assessment levels that your child got?	Yes	No	No answer
	467 (82)	65 (11)	38 (7)

If no, which ones did you agree with? ³	Speaking and listening	Reading	Writing	Maths	Science	None
	16 (25)	26 (40)	26 (40)	19 (29)	19 (29)	7 (11)

Table B.67 - Parent questionnaire – Q6 – Child's agreement with the assessments

¹¹© Qualifications and Curriculum Authority 2004

Did your child agree with the	Yes	No	Don't know	No answer
	338 (59)	24 (4)	158 (28)	50 (9)

Table B.68 - Parent questionnaire – Q7 – Child's concern over receiving the assessments

Was your child worried about receiving the Teacher Assessment levels?	Yes	No	Don't know	No answer
	37 (7)	482 (85)	41 (7)	10 (2)

Table B.69 - Parent questionnaire – Q8 – Parental demand for test results

Did you ask for any of your child's Test levels?	Yes	No	No answer
	82 (14)	475 (83)	17 (2)

If no, why not? ¹²	The Teacher Assessment levels gave me enough information	I don't think the Test Levels are very important	I didn't know I could request the Test Levels	I didn't know the Test Levels existed	Other
	196 (41)	47 (10)	216 (46)	45 (10)	36 (8)

If yes, which ones? ¹³	Reading	Writing	Spelling	Maths
	64 (78)	68 (83)	63 (77)	74 (90)

If yes, why? ¹³	I like to have all available information	I wasn't satisfied with the accuracy of the Teacher Assessment levels	I didn't think the Teacher Assessment levels gave me enough information	Other
	74 (90)	2 (2)	2 (2)	8 (10)

Table B.70 - Parent questionnaire – Q9 – Previous experience of KS1 as a parent

Do you have any other children who have been assessed at Key Stage 1 in previous years?	Yes	No	Don't know	No answer
	246 (43)	301 (53)	16 (3)	7 (1)

If yes, are this year's arrangements better? ¹⁴	Yes	No	Don't know	No answer
	122 (50)	40 (16)	79 (29)	5 (2)

Teacher (1) questionnaire

115 questionnaires analysed.

Note that the first part of this questionnaire (questions 1 to 8) concern the KS1 assessment arrangements in 2002/3, not the trial arrangements of 2003/4.

¹² Multiple selections were allowed here with percentage out of those answering 'No'

¹³ Multiple selections were allowed here with percentage out of those answering 'Yes'

Table B.71 - Teacher (1) questionnaire - Q1 – Amount of work in the pre-trial arrangements

How much work was required in doing the assessments (in 2002/3)?	Minimal	A small amount	A reasonable amount	A large amount	Too much	No answer
	0 (0)	0 (0)	13 (11)	66 (57)	28 (24)	8 (7)

Table B.72 - Teacher (1) questionnaire – Q2 – Levels of pre-trial support

How well-supported were you in terms of information, training and guidance that you received from:	Very poorly	Poorly	Adequately	Well	Very well	No answer
(a) your school?	3 (3)	1 (1)	40 (35)	42 (37)	20 (17)	9 (8)
(b) the LEA?	0 (0)	0 (0)	31 (27)	39 (34)	35 (30)	10 (9)
(c) QCA (exemplar materials etc)?	0 (0)	11 (10)	32 (28)	49 (43)	11 (10)	12 (10)

Table B.73 - Teacher (1) questionnaire – Q3-6 – Views on the pre-trial arrangements

Give your opinion of the following statements concerning the 2002/3 assessment arrangements:	Strongly disagree	Disagree	Agree	Strongly agree	Don't know	No answer
3. They provided consistent and comparable information on children's performance over the year.	4 (4)	44 (38)	57 (50)	3 (3)	0 (0)	7 (6)
4. They provided a good measure of children's attainment levels at the end of the year.	6 (5)	22 (19)	78 (68)	2 (2)	0 (0)	7 (6)
5. They had a positive effect on teaching and learning in the classroom.	38 (33)	49 (43)	18 (16)	0 (0)	3 (3)	7 (6)
6. They gave meaningful information and feedback for parents on their children's progress.	17 (15)	58 (51)	27 (24)	2 (2)	4 (4)	7 (6)

Table B.74 - Headteacher (1) questionnaire – Q7 – Understanding of the pre-trial arrangements

They [the 2002/3 arrangements] were clear and well understood:	Yes	No	Don't know	No answer
(a) by me	106 (92)	2 (2)	0 (0)	7 (6)
(b) by parents	22 (19)	57 (50)	29 (25)	7 (6)
(c) by children	40 (35)	45 (39)	23 (20)	7 (6)

Table B.75 - Teacher (1) questionnaire – Q8 – Using test/task results to inform TA

For how many children did you use test/task results to help you to arrive at TA levels?	All children	No children	Some children	No answer
Reading	51 (44)	27 (23)	26 (23)	11 (10)
Writing	50 (43)	34 (30)	20 (17)	11 (10)
Mathematics	47 (41)	32 (28)	26 (23)	10 (9)

Table B.76 - Teacher (1) questionnaire – Q9 – Comparative impact on workload

In doing the assessments, compare the workload you expect this year against that required last year.	A lot less	A little less	About the same	A little more	A lot more	Don't know	No answer
	3 (3)	46 (40)	35 (30)	13 (11)	11 (10)	4 (4)	3 (3)

Table B.77 - Teacher (1) questionnaire – Q10-13 – Views on the trial arrangements

Rate how strongly you agree overall with the following statements concerning the trial arrangements:	Strongly disagree	Disagree	Agree	Strongly agree	Don't know
10. They will provide consistent and comparable information on children's performance over the year.	1 (1)	14 (12)	84 (73)	13 (11)	3 (3)
11. They will provide a good measure of children's attainment level at the end of the year.	1 (1)	7 (6)	94 (82)	13 (11)	0 (0)
12. They will have a positive effect on teaching and learning in the classroom.	19 (17)	31 (27)	45 (39)	12 (10)	8 (7)
13. They will give meaningful information and feedback for parents on their children's progress.	8 (7)	30 (26)	54 (47)	9 (8)	14 (12)

Table B.78 - Teacher (1) questionnaire – Q14 – Extent of keeping children's work

To what extent are you keeping examples of children's work this year:	Not keeping examples	Keeping one piece per week in each subject	Keeping several examples of work for each child in each area	No answer
(a) in writing?	7 (6)	3 (3)	84 (73)	21 (18)
(b) in mathematics?	7 (6)	4 (3)	81 (70)	23 (20)

Table B.79 - Teacher (1) questionnaire – Q15 – Attendance at LEA training

Have you attended (or will you be attending) your LEA's assessment/moderation training sessions for Year 2 teachers?	Yes, all	Yes, some	No, none	Don't know
	62 (54)	39 (34)	11 (10)	3 (3)

Table B.80 - Teacher (1) questionnaire – Q16 – Teacher understanding of the arrangements

Overall, I am clear about what is involved in the new assessment arrangements.	Yes	No	No answer
	108 (94)	5 (4)	2 (2)

Teacher (2) questionnaire

88 questionnaires analysed.

Table B.81 - Teacher (2) questionnaire - Q1 – Amount of work in the trial arrangements

How much work was required in doing the assessments?	Minimal	A small amount	A reasonable amount	A large amount	Too much	No answer
	2 (2)	0 (0)	40 (45)	34 (39)	10 (11)	2 (2)

Table B.82 - Teacher (2) questionnaire – Q2 – Meeting deadlines

Did you meet all your required deadlines?	Yes	No
	86 (98)	2 (2)

How did this compare to last year?	Easier	About the same	Harder	No answer
	25 (28)	38 (43)	13 (15)	12 (14)

Table B.83 - Teacher (2) questionnaire – Q3 – Comparing teacher workload

In relation to the assessments, compare the workload you experienced this year against that last year.	A lot less	A little less	About the same	A little more	A lot more	Don't know	No answer
	5 (6)	22 (25)	23 (26)	16 (18)	10 (11)	6 (7)	6 (7)

Table B.84 - Teacher (2) questionnaire – Q4 – Comparing teacher assistant workload

In relation to the assessments, compare the workload experienced by teaching assistants this year against that last year.	A lot less	A little less	About the same	A little more	A lot more	No answer
	9 (10)	3 (3)	46 (52)	8 (9)	1 (1)	8 (9)

Table B.85 - Teacher (1) questionnaire – Q5 – Extent of keeping children's work

To what extent did you keep examples of children's work:	Didn't keep examples	Kept one piece per week in each subject	Kept several examples of work for each child in each area	No answer
(a) in writing?	6 (7)	5 (6)	61 (69)	16 (18)
(b) in mathematics?	10 (11)	3 (3)	54 (61)	21 (24)

Table B.86 - Teacher (2) questionnaire – Q7 – Levels of support during the trial

How well-supported were you in terms of information, training and guidance that you received from:	Very poorly	Poorly	Adequately	Well	Very well	Don't know	No answer
(a) your school?	0 (0)	1 (1)	36 (41)	25 (28)	24 (27)	1 (1)	1 (1)
(b) the LEA?	0 (0)	2 (2)	21 (24)	37 (42)	27 (31)	0 (0)	1 (1)
(c) QCA (exemplar materials etc)?	1 (1)	3 (3)	34 (39)	27 (31)	11 (13)	7 (8)	5 (6)

Table B.87 - Teacher (2) questionnaire – Q8 – Number of pupils given specific tests and tasks

Reading	Test/task	2003	2004	Both
	Level 1 task	27 (31)	31 (35)	1 (1)
Level 2 task	21 (24)	34 (39)	3 (3)	
Level 2 test	23 (26)	46 (52)	9 (10)	
Level 3 test	25 (28)	53 (60)	8 (9)	

Writing	Test/task	2003	2004	Both
	Short task	40 (45)	44 (50)	3 (3)
Long task	39 (44)	45 (51)	4 (5)	
Spelling test	25 (28)	60 (68)	2 (2)	

Maths	Test/task	2000	2001	2002	Several
	Level 1 task	3 (3)	15 (17)	33 (38)	1 (1)
		2003	2004	Both	
Level 2 test	28 (32)	53 (60)	6 (7)		
Level 3 test	25 (28)	54 (61)	6 (7)		

Table B.88 - Teacher (2) questionnaire – Q9-12 – Views on the trial arrangements

Give your opinion of the following statements concerning the trial assessment arrangements:	Strongly disagree	Disagree	Agree	Strongly agree	Don't know
9. They provided consistent and comparable information on children's performance over the year.	2 (2)	7 (8)	57 (65)	17 (19)	3 (3)
10. They provided a good measure of children's attainment levels at the end of the year.	1 (1)	6 (7)	58 (66)	20 (23)	3 (3)
11. They had a positive effect on teaching and learning in the classroom.	7 (8)	26 (30)	32 (36)	18 (21)	5 (6)
12. They gave meaningful information and feedback for parents on their children's progress.	6 (7)	21 (24)	40 (46)	16 (18)	5 (6)

Table B.89 - Teacher (2) questionnaire – Q13 – Understanding of the trial arrangements

The trial arrangements were:	Agree	Disagree	Don't know	No answer
12. clear and well understood by me.	86 (98)	2 (2)	0 (0)	0 (0)
13. clear and well understood by parents.	23 (26)	24 (27)	39 (44)	2 (2)
14. clear and well understood by children.	47 (53)	11 (13)	25 (28)	5 (6)

Table B.90 - Teacher (2) questionnaire – Q14 – Variation between test/task results and TA

For how many children was there a variation between the level achieved in the test/task and the final reported level (teacher assessment):	None or almost none	Some	Many	All or almost all	No answer
(a) in reading?	52 (59)	34 (39)	1 (1)	1(1)	0 (0)
(a) in writing?	41 (47)	42 (48)	3 (3)	1 (1)	1 (1)
(b) in mathematics?	50 (57)	32 (36)	5 (6)	1 (1)	0 (0)

Table B.91 - Teacher (2) questionnaire – Q15 – Parental demand for test results

How many parents specifically asked for their child's tests results:	None or almost none	Some	Many	All or almost all	No answer
(a) in reading?	67 (76)	12 (14)	2 (2)	1(1)	6 (7)
(a) in writing?	67 (76)	12 (14)	2 (2)	1(1)	6 (7)
(b) in mathematics?	65 (74)	13 (15)	3 (3)	1(1)	6 (7)

Table B.92 - Teacher (2) questionnaire – Q16 – Levels of teacher confidence

How confident were you in the judgements you had to make when balancing tests, tasks and children's work to produce one overall attainment level for each child:	Not at all confident	Not very confident	Quite confident	Very confident	Don't know	No answer
(a) in reading?	0 (0)	5(6)	38 (43)	43 (49)	1 (1)	1 (1)
(a) in writing?	1 (1)	6 (7)	38 (43)	41 (47)	1 (1)	1 (1)
(b) in mathematics?	1 (1)	9 (10)	34 (39)	42 (48)	1 (1)	1 (1)

Table B.93 - Teacher (2) questionnaire – Q17 – Views on roll-out

Do you think that the trial arrangements should be made permanent from 2005?	Yes	No	Don't know	No answer
	72 (82)	7 (8)	7 (8)	2 (2)

Interview schedules

These are given in the following order:

- C.1 LEA coordinator interview schedule
 - C.2 Interview Schedule for Teachers in evaluation schools
 - C.3 Interview Schedule for Teachers in non-evaluation schools
- (The last two of these only differ in their last two questions)

LEA coordinator interview schedule

Duration of interview 30 minutes

1. Your workload	How was the workload for you and your colleagues compared to last year? Do you expect that to change next year?
2. Schools	What seem to be the benefits to schools, so far? ols have found most challenging in the new arrangements? Do these benefits and/or challenges apply to particular types of schools, or across the board?
3. Workload in schools	What is your impression of the workload issue in schools? Is the same level of supply cover being offered in schools, do you think?
4. School management issues	Have you come across any school management issues other than workload?
5. Teacher judgements	Is it your impression that teachers are tending to just use the tests and tasks and report the outcomes as teacher assessments, or do they have more confidence to draw on the full range of evidence when making their own judgements? How do you think any differences between tests results and teacher judgements are being resolved, typically?
6. Moderation	How has the moderation system you have been using worked out in practice? What have you learned from it? What would you do differently in future? Have the changes affected your relationships with schools? Was it more work for you or less?
7. Substitute results	As it turns out, have you needed to substitute LEA assessments for school's judgements anywhere? How does that compare to previous years?

8. Other issues	What other outstanding issues are there for the LEA in these new assessment arrangements? Do you think the approach used in the trial should be ‘rolled out’ to all LEAs next year? If it is, what needs to happen for it to be successful?
------------------------	--

Interview Schedule for Teachers in evaluation schools

Duration of interview: 20 minutes

<p>1. Opener</p>	<p><i>Overall how did you find the arrangements for assessment this year, compared to last year?</i> (Prompts: Straightforward? Stressful?)</p>
<p>2. Flexibility of arrangements</p>	<p>Did you like the flexible arrangements for test and task administration? Do you feel you were able to demonstrate greater professional judgement? Did you like the flexibility of being able to use 2003 tests and tasks?</p>
<p>3. Record keeping</p>	<p>Have you approached keeping records differently this year from last? What have you done differently? What kinds of new records did you keep? (for reading, writing and mathematics) Did you keep portfolios of children's work this year? Is this something you've done before?</p>
<p>4. Differences in test / TA</p>	<p>When the test results were different from what you expected for a particular pupil, how did you resolve this in practice? For roughly what proportion of the children was this the case? Did you administer fewer tests overall this year? To what extent? Did you know the 2003 tests are in the public domain? And if so, did this affect your plans?</p>
<p>5. Accuracy</p>	<p>In general, do you think the assessment arrangements this year have resulted in the same level of accuracy as last year? (Prompts: more accurate? less accurate?)</p>
<p>6. Effect on pupils</p>	<p>Do you feel the arrangements have been better for your pupils this year compared to last? Did they respond differently to the tests and tasks? Did you prepare pupils for the tests differently compared to last year? Did you present the tests and tasks differently because they were not going to be reported?</p>
<p>7. Effect on Teaching & Learning</p>	<p>Do you think the pilot had any effect on teaching and learning in the classroom? Did you learn more about the pupils' understanding and knowledge? Did the different arrangements affect your approach to the marking of the tests and tasks? Were you less strict about the mark scheme?</p>
<p>8. Adjusting to the new arrangements</p>	<p>Overall, how much time and effort was taken in adjusting to the new system? Was much of the effort transitional so that once you were familiar with what was required any additional workload largely eased?</p>

Interview schedule for teachers in non-evaluation schools

Duration of interview: 20 minutes

<p>1. Opener</p>	<p>Overall how did you find the arrangements for assessment this year, compared to last year? (Prompts: Straightforward? Stressful?)</p>
<p>2. Flexibility of arrangements</p>	<p>Did you like the flexible arrangements for test and task administration? Do you feel you were able to demonstrate greater professional judgement? Did you like the flexibility of being able to use 2003 tests and tasks?</p>
<p>3. Record keeping</p>	<p>Have you approached keeping records differently this year from last? What have you done differently? What kinds of new records did you keep? (for reading, writing and mathematics) Did you keep portfolios of children's work this year? Is this something you've done before?</p>
<p>4. Differences in test / TA</p>	<p>When the test results were different from what you expected for a particular pupil, how did you resolve this in practice? For roughly what proportion of the children was this the case? Did you administer fewer tests overall this year? To what extent? Did you know the 2003 tests are in the public domain? And if so, did this affect your plans?</p>
<p>5. Accuracy</p>	<p>In general, do you think the assessment arrangements this year have resulted in the same level of accuracy as last year? (Prompts: more accurate? less accurate?)</p>
<p>6. Effect on pupils</p>	<p>Do you feel the arrangements have been better for your pupils this year compared to last? Did they respond differently to the tests and tasks? Did you prepare pupils for the tests differently compared to last year? Did you present the tests and tasks differently because they were not going to be reported?</p>
<p>7. Workload</p>	<p>How was your workload when doing the assessments compared to last year? What was the most time consuming element of this? (Prompts: Tests / tasks / coming to a final teacher assessment; reading / writing / mathematics) Have the deadlines presented any difficulties for you? Was the assessment workload spread over time or very concentrated into short periods? <i>What about the workload of others in the school, such as teaching assistants, the school secretary, the head?</i></p>

8. Training	What training did you receive about the new arrangements for Key Stage 1 assessment? How effective and worthwhile did you find it? What written support material did you find useful? Were you given the QCA exemplification materials? Did you use them? If so, how useful did you find them (very / quite / not very / not at all)
--------------------	--

Focus group schedules

These are given in the following order:

D.1 Parent focus group schedule

D.2 Year 2 teacher focus group schedule

Parent focus group schedule

1. [The facilitator's opening statement]:

My name is ... I am one of the team based in Leeds who are conducting the independent evaluation of the new school assessment arrangements for the end of Key Stage 1.

The purpose of this meeting is to find out what parents generally feel about the changing arrangements. We are also sending out questionnaires, but meetings like this can go into more depth than questionnaires can.

In case you don't all recognise each other, the group are all parents from the school who have volunteered to take part [but also mentioning any variations from this].

Please would you write your first name on one of these labels and put it on your lapel.

You will notice that there is a minidisk recorder in the centre. We need to make a recording so that we are sure we collect all the opinions voiced, but we will not make any use of the recording other than to make a paper record – a transcription – and the audio recording will then be destroyed. Also, all the names that are said today will be removed from the paper version. For example, if you say the name of your child, we will put my child instead, or you name your child's teacher we will change it to my child's teacher. Is that OK?

The meeting is not to decide anything but to get at the range of views that parents have. You can and should express different opinions, if genuinely felt.

My role in the meeting is to keep it focused on the key topic, keep it moving on to cover a range of issues, and to ensure it finishes on time. I am not supposed to express any views of my own, so if I do, stop me.

For most of the meeting we will talk as a whole group, but for this first bit I am going to ask you to talk in smaller groups.

2. [The 'ice-breaker']

Talk for a couple of minutes with the person next to you about this question:

What do most parents want from school assessment of their children?

Then decide between you on what you think are the most important things – up to three - and write them on the sheet.

[Question on a sheet with three bullet points that is distributed, with a pen, and a

© Qualifications and Curriculum Authority 2004

board to write on if there is no table.]

[Reporting back by going round and round asking for different ideas from each group in turn, until all ideas have been aired, then collect in the sheets.]

3. [Information]

Now, I'd like to ask about the information you've had about the assessment arrangements for your child at the end of Key Stage 1. Overall, do you feel that you've been well informed? For example, have you been given any leaflets on it or has your child's teacher spoken to you about it?

So, do you feel you know enough about what the arrangements have been?

[Opinions round the group]

Here is a reminder of the assessment arrangements, and how they differ from what was done before.

[Distribute copies of the summary sheet, and then read through it and explain anything that parents need further explanation of.]

4. [Publishing of TA / task & test results]

What you will notice from the information sheet is that the results you are given as parents have changed. In this new system, you get Teacher Assessment (TA) levels that have been based both on the test and task results and on teacher's observations of children working through the year, whereas before you would have got the levels based on teachers observations separately from the test and task results.

So my question is does this change worry you, in that you don't automatically get the test and task results for your child?

[Opinions from the group.]

You can still get the separate test and task results from the school, but only by asking for them, since they're not automatically sent out. If I could just ask then, are you interested in asking for these test and task results? And indeed have any of you asked for them?

[Opinions from the group]

5. [Confidence in reported TA]

Now I'd like us to talk about the actual Teacher Assessments (TA) that you have been given by the school for your child, on a sheet like this.

[‘Wave’ at them a blank copy of the sheet that the parents get from the school informing them of their child's levels]

There would have been a level of W, 1, 2 and so on for each of the main subjects.

© Qualifications and Curriculum Authority 2004

My question is do you have confidence in these Teacher Assessment levels?

[Opinions from the group]

Do you think the Teacher Assessment levels are accurate?

[Opinions from the group]

Would it affect your confidence in the Teacher Assessment levels if you were also given the test and task results at the same time?

[Opinions from the group]

6 [Effect on children]

Moving on now from the results given to you about your child, to looking at how you feel he or she has handled the whole experience of being assessed and having the results sent home.

Do you think your child found it stressful, or did they just take it in their stride and didn't seem to be that bothered or were even perhaps unaware of what was going on?

[Opinions from the group]

Did they handle it the way you expected them to?

[Opinions from the group]

7 [Overall opinion]

So, finally, what about these new arrangements? Is this what is needed?

[Opinions from the group]

8 [Wind up time]

That is all. Thank you for your time. Does anyone have any other comment to make, that they feel they have not yet had the opportunity to say?

Year 2 teacher focus group schedule

1. *Please discuss with one or two people sitting near you, with one of you willing to report back, what you have found good and not so good about the new arrangements.*

(Reporting back: What have you liked about the new arrangements?

What have you felt have been troublesome aspects?)

2. [Unless covered by answers to the previous question.]

One of the supposed benefits of the new arrangements is the flexibility - to do tests or tasks of your choosing from some options, at a time that suited you. Have you been able to take advantage of this? And has it been a help?

3. *Do you think the assessment levels you have reported for each child are more accurate this year than in previous years?*

4. *Were the children less affected by the arrangements for assessment than in previous years?*

How noticeable has been any pressure to perform well from the school or from parents?

5. *Do you think that these arrangements have involved more work than the old arrangements?*

Which staff - teachers, teaching assistants, secretary, head teacher, management team, have been most affected in terms of changing workload?

[If workload has increased] *Do you think that if the new arrangements continue, the workload will always be more, or has it been more because of learning a new system? [and being involved in the evaluation, if that applies]*

6. *How pressured did you feel to do the assessments in a particular way, for example to use a particular task or test, to choose a task or test result over your teacher assessment judgement, or to award as high a level as possible?*

[If pressure was felt] *Where was the pressure coming from?*

7. *How interested have parents been in the assessment arrangements?*

8. *Is anyone in an infant or first school where the results have been sent up to the juniors? How have they reacted to the new system?*

9. And finally, is it worth it? Should this new arrangement be rolled out to all schools next year?

Moderation plan summaries

Table E.1 Moderation plans for the 10 LEAs in the evaluation

LEA	No. KS1 schools	No. schools moderated-full audit	Choice of schools	Moderation visits planned for	Special
1	53	14-20	Schools causing concern, due for audit	Beginning of June	
2	39	10	due for audit	Late in summer term	
3	82	21 / 25	Knowledge of school, history of KS1, geographical spread, size/type, new Y2 teachers	mid June	Teacher preselects 3 or 4 pupils who can be the focus for the moderator to observe classroom working.
4	109 + 6 special	100%	All chosen	From 17 May at the earliest & through to June	
5	51	15	14 due for moderation and 1 school chosen due to concerns which arose during last year's moderation procedures	W/c 14 June	
6			No form submitted		
7	128	25% + schools of concern	Schools opt in & then selected from the 25% due for audit	May	
8	228	56	Schools causing concern, teachers new to Y2, geo. location, size/type	In March	Moderation will be for 3 pupils in each Y2 class; a higher attaining pupil, an average attaining pupil' and a lower attaining pupil.
9	416	104 + others if concern arises	due for audit	during June	Teacher asked to complete form & moderator preselects children
10	437	25%	due for audit	Across spring & summer terms	Only 12% of schools have one to one visit, to watch tests/tasks, look at work etc. 12% get small group (4 schools) meetings with a moderator

Evaluation of the trial assessment arrangements for key stage 1

Table E.2 – Moderation plans for the LEAs in the trial but not in the evaluation

LEA	No. KS1 schools	No. schools moderated- full audit	Choice of schools	Moderation visits planned for
11	295	75	Due for audit + those where assessments were of concern in 2003	Second half of spring term or early summer term
12	69	20	Random, but all schools audited in four year period + schools with issues moderated more.	tba
13	230	70	Those due for audit, infant only schools, schools where problems were presented in past, NQT only Y2 teacher	
14	86	30	Those due for audit + those causing concern	Spring
15	307	120	Identified by LEA though monitoring processes + those with serious errors in previous year's audit	First two weeks in June
16	224	61	Those due for audit	May/June
17	52	14	School request, those due for audit, advisor/moderator request	2 visits per school: 1 May & 1 June
18	No form returned			
19	29	8	Those due for audit, assessment known to be weak, new or inexperienced teachers	2 visits: 1 April/May & 1 early June
20	140	36	Those due for audit	Early June
21	69	24	Those due for audit, with issues arising from last year's audit, in special measures, with serious weakness or without experienced Y2 teacher	June
22	133	33	Those due for audit + any with concerns	May/June
23	46	12	Those due for audit & any schools where issues were raised during 2003 audit	May/June
24	58	15-17	Those due for audit & any schools where issues were raised during 2003 audit	May/June
25	58	16	Those due for audit & schools with issues from previous year	May/June
26	245	73	Those due for audit & 10 schools where there is cause for concern	One visit to every Y2 teacher to observe a test or task
27	36	11	Those due for audit + schools identified from previous audit	Subject to discussion with Headteachers
28	131	32	Those due for audit, range of socio-economic contexts, across Sheffield though LEA cluster system	W/c 21 June
29	29 + 4 special	8	Those due for audit & random visit	Beginning of June
30	300	25%	Those due for audit + serious weaknesses, special measures, concerns, new to year	May/June
31	66	16	Those due for audit + others because of concerns raised by LEA, school or QCA	June
32	108	27	Those due for audit + any schools with OFSTED or Primary Advisor concerns, + teachers new to Y2	June
33	52	14	Those due for audit + some because of concerns raised by LEA, school, QCA	June
34	208	25%	Schools with significant level 3 marking inaccuracies in 2003, schools not subject to audit last year, randomly selected schools	To be agreed with school