![Qualifications and Curriculum Authority logo]

**Qualifications and
Curriculum Authority**

Evaluation of the assessing pupils' progress in key stage 2 pilot project 2006–2008

Final report

Contents

Please note that all quotes used in this report came from teachers and headteachers unless otherwise attributed.

Executive summary

This is the final report from the two-year pilot assessing pupils' progress (APP) at key stage 2 project undertaken by the Qualifications and Curriculum Authority (QCA) in partnership with the National Strategies and funded by the Department for Children, Schools and Families (DCSF).

This report refers to some features that were specific to the pilot phase and have informed the wider development of APP. These features will not apply to the APP approach when used by schools beyond the pilot (see annex on page 154).

The project explored, through work with pilot schools and their local authorities, two key questions:

- Does APP provide an effective model for teacher assessment within key stage 2?

- Is the APP assessment model manageable?

In agreeing to take part in this pilot, teachers and headteachers were hoping to improve the quality of pupil assessment in their schools. Feedback over the two years of the project has demonstrated that teachers, headteachers and local authorities believed that APP can and does improve teacher assessment.

> 'APP has provided an accurate method of [assessing] pupils' progress which has given teachers an in-depth understanding/knowledge of their progress and the whole child, rather than the snapshot that testing gives.'

Feedback from local authorities was that APP encouraged a shared ownership of responsibility for achievement and progress compared to a view which they felt had existed in some schools that it was the year 6 teacher at the end of the key stage who was responsible for making sure that pupils achieved their potential. Moreover, teachers valued this ownership:

> 'From just six pupils, it highlighted to me areas of weakness, not only in my year group but the year group below too.'
>
> 'We have a raised profile and credibility.'
>
> 'Teachers feel their professionalism is trusted.' (local authority consultant)

This sense of the importance and value being placed on teacher expertise meant that, even when coping with the inevitable challenges and difficulties, the pilot activities organised over the course of two years have been conducted in a generally positive and, in some cases, extremely enthusiastic atmosphere. Obviously there has been a variety of ways in which individual schools and teachers have responded but for a number of schools APP has provided a catalyst for extensive change and development.

> 'It has triggered a full assessment review in our school.'
>
> 'I have never been associated with anything that has been so effective.' (Moderator)

'Overall process is fantastic.'

Outcomes from the experience of the pilot

The following are the key observations based on the evidence generated from the pilot evaluation that are most helpful in promoting the effective use of APP in a way which is manageable for teachers.

Introducing APP for the first time

Emphasis needs to be on the nature of APP as a periodic assessment, with teachers being given an understanding of the differences between periodic and continuous assessment. The purpose of the assessment focuses needs to be explained and how they differ from learning objectives. Whatever form initial training takes, it needs to allow teachers to actively engage with the use of the assessment guidelines and other APP materials in a real and relevant context. Teachers found their first experience of trying to apply the guidelines to the evidence in their own pupils' work much more useful in helping them get to grips with the assessment focuses than a 'theoretical' training session. Through feedback from a variety of different sources, recommendations have been made that teachers new to APP should be introduced to the benefits and challenges by other teachers currently using APP in their own classrooms. The provision of high quality training was seen as reinforcing the status of APP as a credible and powerful approach to teacher assessment:

> 'It is so valuable all schools should be offered the chance to attend free training. I keep coming across local schools who know nothing about it. If this is going to be a major national initiative that makes a difference, it needs to be pursed by the local authorities with more drive.'

Time

Initially it is important to allow time for teachers to develop familiarity and expertise – to practise their APP skills. This can be supported by a phased approach when taking on APP, for example starting with a small number of pupils, testing judgements informally with more experienced colleagues, or focusing on a single subject or attainment target. As processes become more established, time will still be needed to maintain the consistency and quality of judgements and to make sure that key messages from assessment are incorporated into planning. From feedback it seems that the structure that schools find most effective is a termly cycle, which brings together standardisation, assessments, moderation and planning review. Most importantly, teachers need the opportunity to work with colleagues in their own school and in other schools to help in implementing improvements to practice in all aspects of assessment and to give them confidence in their own judgements and those of their peers.

Securing consistent assessment standards

Teachers and headteachers want assessment judgements to be reliable and defensible. They have valued the opportunities in the pilot to endorse APP judgements through moderation. The report on the pilot of the various moderation models has considered ways in which consistency could be maintained across schools and different local authorities. However, moderation of APP judgements

has to be clearly rooted in the principles of APP assessment, valuing teachers' professionalism and taking account of evidence of achievement from every available source without overemphasis on pupils' written work. Innovative approaches to sharing information should be considered so that successful moderation does not rely on wasteful processes like photocopying. Where possible the lessons learned during a moderation should be followed up in school so that maximum impact of the training elements of moderation is assured. In the pilot, moderation has been very frequent but outside the pilot fewer moderation activities would be expected, so it will be more important to capitalise on the valuable training opportunities they offer.

Modelling whole class and whole school implementation

Within the pilot there has been no 'directive' guidance to teachers or schools about how to move from the initial requirement for the assessment of a focus group of pupils to a whole school implementation. We know from the evaluation that many schools appear to have done this successfully. This has been achieved in the context of the pilot with support available from local authorities in terms of meetings, follow up visits to schools and communication with the project team at QCA. Outside the pilot, as schools have the opportunity to pick up and use APP, the support available to schools should include as much help as possible on how to manage the transition to a full implementation of APP assessment.

Support for schools

There has been considerable variation within the pilot in the nature of the involvement of senior leaders in schools and also in the support offered by the pilot local authorities. Messages from the evaluation are quite definite. Those schools which entered the pilot with senior leaders not fully committed or in local authorities not offering coherent support for teachers working together have been the most likely to drop APP or to limit use to the original 'two teachers' requirement. The support offered by the proactive local authorities has been valued highly by teachers. Drop-in support and advice sessions, ideas and materials for activities which produce good assessment evidence have been welcomed. These local authorities have been able to build a nucleus of assessment expertise within key stage 2, growing a resource to develop and support moderation and training within their areas. Investment in helping teachers develop to work at all levels in securing sound and reliable assessments sends a message that processes which promote achievement and progress are owned by all teachers.

1. Introduction

The assessing pupils' progress (APP) pilot project was organised by QCA with the support of the National Strategies to trial a new approach to teacher assessment of English (reading and writing) and mathematics in primary schools. (Prior to autumn 2007, the pilot was known as the monitoring children's progress project).

This approach is intended to support teachers in developing their skills in assessing progress in reading, writing and mathematics within the context of a broad curriculum. APP promotes a model of assessment which is the periodic, systematic review of achievement as evidenced through a range of sources, including what teachers 'know' about their pupils as a result of everyday classroom interactions. Evidence is reviewed using structured assessment criteria with two outcomes:

- a profile of strengths and weaknesses across a range of assessment focuses in each attainment target to help determine next steps in teaching and learning

- a judgement expressed as a national curriculum level for reading, writing and for each mathematics attainment target.

The intention is that assessments are derived from a broad evidence base, rather than being dependent on the outcomes of a 'one-off' task or test.

When this pilot began in autumn 2006, related development work had already been undertaken within key stage 3 and some limited development and trialling of materials for key stage 2 had been run with a group of year 5 pupils in seven local authorities in 2005-2006. The full key stage 2 pilot ran for two years until July 2008. This report brings together the findings from a range of evaluation activities commissioned to fulfil two purposes:

- to investigate the effectiveness and manageability of APP in offering a model for assessment of reading, writing and mathematics within key stage 2

- to provide immediate and regular feedback to the project team on emerging issues for schools and local authorities to support the work within the pilot.

Interim reports have been written for the project team and informal feedback provided at frequent intervals to inform project planning and the further development of materials.

Sources for this report

This report draws on the following:

- responses to questionnaires from teachers at five time points. The specific focus for each questionnaire varied with the stage of the project:

    1. Autumn 2006 – teachers' first experience of using APP

    2. Spring 2007 – using outcomes from APP and the effect of increasing familiarity

    3. Summer 2007 – their experience in the first year and plans for second year

    4. Autumn 2007 – the organisation and challenges of second year of pilot

    5. Summer 2008 – embedding APP in school practice and the impact of the pilot

- responses to a questionnaire for headteachers at the end of the first year of the pilot

- completed evaluation forms from training and moderation events

- feedback reports to individual schools prepared by external moderators

- observation of a range of local authority review meetings

- outcomes of moderation 'accuracy' checks and other feedback from moderators

- attendance at moderation events across the two-year pilot

- analysis of assessment outcomes data collected from participating schools at specified time points in each of the six terms in the pilot.

While all the interim reports have had as their main focus the key questions of effectiveness and manageability, each has also included specific issues for the project team to consider for the next steps in the project. For example, the reports have made suggestions aiming to inform moderator training or communication with schools and local authorities.

Appendix 1 includes all the questionnaires used within the evaluation. In the first year of the pilot the intention was to offer each participating teacher an opportunity to respond to the questionnaires if they wished. In the second year, the intention was that one respondent from each school should provide a collective view from their school.

Organisation of the project

The work of the project team at QCA was overseen by a steering group with members from both QCA and the National Strategies. A core project team at QCA developed materials and recruited participating local authorities. As the project developed over two years, organisational changes were made which were significant in terms of the teachers' experiences of APP and the outcomes from the evaluation.

First year

Twelve local authorities originally volunteered to be part of the project with 104 schools taking part. Headteachers were introduced to the project through a series of briefing meetings and these were followed by two full training days for nominated teachers, led by the QCA project team and held in late November and early December. Participating schools were required to trial APP with pupils working between level 2 and level 4 in English or mathematics, or in both subjects. Schools were encouraged to make the decision as to which year groups, pupils, teachers and subjects would be most appropriate to their own needs. Attendance at the initial training session was limited to two teachers per school per subject. Schools then had the option to cascade training to additional teachers if required. The focus was primarily intended to be on pupils in year 3, 4 or 5 but some schools were anxious to include pupils and teachers in year 6. As a result there was a small cohort of year 6 pupils in the first year of the pilot but their participation was on a restricted basis, since level 5 assessment criteria were not available until the summer term of the first year of the project.

At the start of the project, schools were given very specific instructions about the way in which APP assessment should be organised. Each school had to produce a plan setting out activities and time-lines and naming key members of staff which they then shared with their external moderator. Before completing the assessment guidelines for pupils, teachers were expected to undertake an in-school standardisation activity using a model and materials provided during the central training sessions. This was intended to ensure that all teachers had the same appreciation of the standards for each level and a shared understanding of the process. Teachers were asked to use the APP materials with a small focus group of pupils in their class (the suggestion was for six pupils per subject) and carry out three 'rounds' of APP assessments over the course of the year (see annex on page 154). To support the needs of the evaluation, defined time periods were set for these assessments and teachers were required to send off their assessment outcomes to the data collection agency appointed by QCA. Data collection of the assessment outcomes involved the transfer of levels awarded to a simplified version of the guidelines forms, leaving the more informative full guidelines sheets in school to support the next steps in teaching and learning for the focus pupils. Following the completion of their assessments, an in-school moderation exercise was expected. All these processes were modelled during the central training. While this externally imposed 'time frame' was necessary to meet the needs of the project, it was a constraint for the schools because they could not always time their APP assessments to align with their own needs or with existing assessment practice in school. It also meant that one of the key features of APP – that it can be used flexibly as

and when required by teachers – was not always readily apparent to teachers at the start of the pilot.

The QCA team trained a team of external moderators to provide a resource for participating teachers in helping to establish consistency in APP judgements (see annex on page 154). A moderator visit was organised for every school in each of the three assessment time periods, with separate visits focusing on English and mathematics. The focus for each moderation visit was agreed between the teachers and their moderators in advance. It might be on a reading, writing or mathematics attainment target or occasionally on specific assessment focuses within reading or writing. Verbal feedback was given on each occasion by the moderator and this was followed up by a brief written report back to the school.

Some local authorities opted to provide additional support for their teachers in the form of twilight review meetings. The purpose of these varied, with some focusing on gathering feedback and sharing ideas, while in others a more definite training element was included, intended to address issues which the teachers had identified, for example finding evidence to support the assessment of Ma1. In other local authorities, support took the form of visits to schools or simply maintaining contact.

The pilot schools received funding to support the pilot activity, equivalent to five days of supply cover per subject. This was not intended to cover the time required for 'assessment' but as a contribution to time out of the classroom for participation in training, external moderation and evaluation activities.

Second year

The basic requirements of the pilot were maintained in the second year (three rounds of assessment, the provision of APP materials for all schools taking part, collection of assessment outcomes for analysis in each round and the use of a focus group of pupils) but the organisation of the project was far less centralised and allowed for more flexibility of approach at school and local authority levels.

The role of the local authorities in the second year was far more prominent. An early full-day briefing was organised for local authority staff by QCA. There was no central training for teachers from the QCA project team, even where schools were introducing new teachers to APP. Any training and support outside the schools was provided through the local authority. Schools were invited to submit assessment outcomes for pupils beyond a small focus group if they so wished. (In fact most chose to limit the data collection sample to an average of six pupils per subject per year group, but pupils from 'new' year groups were often added at the start of the second year.) They were also offered a much wider 'assessment period window' to allow them to carry out their APP assessments at times which were more convenient for them and which allowed them to take account of local timings for such features as moderation activities.

The QCA project team made changes to the project materials over summer 2007 as a result of the feedback from teachers in the first year. The most important of these addressed concerns about the language for criteria in level 3, which many teachers had found to be negative. Updated versions of the assessment guidelines, flowcharts and standards files were provided to each of the local authorities for distribution to their schools at the start of the autumn term 2007.

The context in the second year was different because of the growing interest in APP and the decision to make the materials available to all schools through a national development programme supported by the Primary National Strategy from autumn 2008. (See http://www.standards.dfes.gov.uk/primaryframework/assessment/app)

Related development and trialling work had implications for the participating schools.

1.   Running alongside the main key stage 2 pilot work was a related project exploring the use of various approaches to the moderation of APP teacher assessment. Local authorities were invited to select a model of moderation for use with their participating teachers. For a very few schools, moderation still took the form of individual visits from their moderator but, for most, moderation in the second year took the form of joint moderation with colleagues in other local schools. A separate report has been produced on the trial of these models. As with the first year, moderation in whatever form always included written feedback to teachers and this feedback continued to offer insights into the effectiveness and manageability of the whole APP process.

2.   In the case of four of the local authorities, APP was introduced into schools beyond the original pilot group. Generally, these schools did not take part in the data collection but were often working beside the pilot schools at moderation, training and support meetings.

3.   In autumn 2007 the DCSF launched its initiative, 'Making Good Progress' (MGP), which encourages the use of periodic teacher assessment, including APP, to inform test entry. One of the local authorities from the APP key stage 2 pilot was involved in the MGP pilot and so their schools from the key stage 2 pilot had a rather different experience in the second year than that of schools in other local authorities. 'APP + MGP' schools and 'MGP' only schools worked together in this local authority at training events and moderation meetings.

4.   A separate pilot project looking at the use of APP within key stage 1 began in autumn 2007. Around half of the schools taking part were also part of the key stage 2 pilot. For the key stage 1 schools there was one day of central training for nominated teachers and their headteachers. One of the local authorities introduced to APP for the first time through the key stage 1 pilot opted to involve partner Junior and Infant schools, working together. These Junior schools did not contribute to the data collection of assessment outcomes in the second year of the pilot but they had an opportunity to feed back comments and experiences into this evaluation.

The level of support within local authorities varied as in the first year, and there were five instances where schools received only limited support from their local authorities for the project activities. As might be expected, the schools from these local authorities were more likely to drop out of the project than those where the local authority took a strong leadership role in the second year.

Some schools, which may have found the collection of assessment data rather burdensome, did not submit data for the second year, even though they took part in moderation activities and had obviously continued the use of APP in school. Seventy-four schools continued to contribute to the termly data collections throughout the second year of the pilot. Funding for pilot schools in the second year was limited to the equivalent of three days of supply cover.

The organisational changes made in the second year of the pilot meant that teacher experiences of APP were perhaps more indicative of what might happen as APP is adopted on a voluntary basis by local authorities or groups of schools. There was more variation in approach to suit local needs, and greater access to local resources and support to supplement the nationally published materials and guidance. At any one time, teachers involved had different levels of familiarity with and experience of the APP process. This provided a useful opportunity to identify some issues for local authorities in taking APP forward in the 'non-project' context. From the point of view of the evaluation, however, it makes it rather more difficult to draw any conclusions about the effect of embedding APP over time, since for many teachers, the second year of the project was in fact their first year of using the materials and many pupils were included in one year of the data collection only. Even those pupils for whom assessment data was collected over two years were not necessarily working in classrooms where APP was in use for the full two years of the pilot, as they may not have remained with teachers who were taking part in the trial. These limitations have been taken into account when reviewing the available sources of evidence.

APP materials and the assessment process

In APP the periodic assessment of pupils' progress involves the review of a range of evidence against criteria for each attainment target in English and mathematics. Within each attainment target a set of assessment focuses (AFs) based on the national curriculum programmes of study have been derived to support the systematic assessment of developing skills. These are fully described in the project documentation but for reference they are set out below. When making assessments, teachers use assessment guidelines laid out in the form of grids which illustrate performance at adjacent levels in each of the assessment focuses. The grids as used within the second year of the project are included in appendix 2. A level judgement is made for each individual assessment focus and teachers then follow a flowchart to arrive at an overall level judgement for the attainment target. To support their work in the project, teachers received a file of materials that explain the process and includes a collection of annotated pupil work that exemplifies performance at a range of levels (the standards files).

Table 1.1: English assessment focuses

Reading	
AF1	Use a range of strategies, including accurate decoding of text, to read for meaning
AF2	Understand, describe, select or retrieve information, events or ideas from texts and use quotation and reference to text
AF3	Deduce, infer or interpret information, events or ideas from texts
AF4	Identify and comment on the structure and organisation of texts, including grammatical and presentational features at text level
AF5	Explain and comment on writers' use of language, including grammatical and literary features at word and sentence level
AF6	Identify and comment on writers' purposes and viewpoints, and the overall effect of the text on the reader
AF7	Relate texts to their social, cultural and historical contexts and literary traditions
Writing	
AF1	Write imaginative, interesting and thoughtful texts
AF2	Produce texts which are appropriate to task, reader and purpose
AF3	Organise and present whole texts effectively, sequencing and structuring information, ideas and events
AF4	Construct paragraphs and use cohesion within and between paragraphs
AF5	Vary sentences for clarity, purpose and effect
AF6	Write with technical accuracy of syntax and punctuation in phrases, clauses and sentences
AF7	Select appropriate and effective vocabulary
AF8	Use correct spelling

Table 1.2: Mathematics assessment focuses

Mathematics Ma1: Using and applying mathematics
Problem solving
Communicating
Reasoning
Mathematics Ma2: Numbers
Numbers and the number system
Fractions, decimals, percentages and ratio
Operations and relationships between them
Mental methods
Solving numerical problems
Written and calculator methods
Mathematics Ma3: Shape, space and measures
Properties of shape
Properties of position and movement
Measures
Mathematics Ma4: Handling data
Processing
Representing
Interpreting

The main section of this report explores evidence from all the evaluation activities organised around issues related to the effectiveness and manageability of APP. The final section brings together key messages and lessons learned from the pilot and examines their implication for the use of APP in future.

2. Introducing the APP criteria and materials

Initial training

Teachers in the pilot were introduced to APP through two separate one-day training sessions. The training sessions included an introduction to the rationale for APP assessment, the materials, using evidence to support APP assessments and the requirements for the pilot in terms of data collection and moderator visits, as well as activities designed to promote understanding of the national standards for levels in KS2 and how these were exemplified in the standards files. This was a substantial input for the time available and it was obvious from immediate feedback at the end of the first training session that teachers were rather overwhelmed and that they would have welcomed more 'hands-on' activities to allow them to become used to the materials. These comments were taken on by the project team and plans for the second sessions were adapted accordingly. Feedback after the second session indicated that most teachers were reasonably confident about their next steps but there were still a substantial minority of teachers who were unclear about aspects of the work.

Observation of the introductory training sessions for key stage 1 in autumn 2007 suggested that significant work had been done by the project team in developing effective ways to introduce the key principles of APP to 'new' teachers, presumably as a result of lessons learned from the key stage 2 pilot. For some teachers in the key stage 2 pilot, even those who were quite confident and who went on to make very effective use of APP, 'mixed messages' remained a barrier which had to be overcome. The two most significant areas of misunderstanding for this group seemed to be in the concept of 'periodic assessment' and how this differed from 'continuous assessment' and the extent to which evidence had to be 'recorded' in order to be valid. These issues will be picked up later in the report. A clear understanding of what APP is 'all about' gives teachers a head start in developing expertise.

Developing familiarity with the assessment focuses

The assessment focuses for English used within APP were developed in the context of national curriculum assessment in 2003 but at the initial training sessions it was clear that generally teachers were unfamiliar with them. Assessment focuses for mathematics were developed to support APP assessment and these were therefore new to all teachers at the start of the project. In mathematics, the content and rationale of the level descriptors also appeared to be unfamiliar to many teachers. Overall it was evident that teachers generally 'understood' performance at particular national curriculum levels in terms of test outcomes, derived by applying formulas in mark schemes or from the use of published assessment schemes with highly atomistic criteria, often applied to single pieces of work. Some were used to analysing test questions and were familiar with the concept of questions designed to elicit a response at a particular level. For some, the idea of a 'spiky' profile of achievement, where an individual could be working at different levels in different aspects of mathematics, reading or writing, was initially quite challenging.

This difficulty in accepting that there could be real, significant variations in achievement across assessment focuses was occasionally evident in teachers new to APP throughout the project. Moderators reported situations where level judgements were based on an expectation of what a 'level 2' pupil should be able to do, rather than on any actual evidence for the particular pupil. This attitude was also seen in a reluctance to look towards the criteria for the level above for the higher achieving pupils. The most immediately obvious impact of APP was in teachers' understanding that two pupils can both be working at the same level and yet have very different profiles across the assessment focuses and therefore very different needs for progression.

From the very first round of assessments, teachers were convinced that the use of APP had given them a better understanding of the characteristics of performance at the levels in which their pupils were working. Chart 2.1 shows their responses from the first questionnaire completed in January 2007, when 92 per cent of teachers considered that APP had already improved their understanding of national curriculum levels. This conviction was reiterated at all the feedback and review meetings and in the final questionnaire to project teachers in summer 2008, when again 92 per cent of the respondents confirmed that this had been a key impact of APP.

Chart 2.1 Teacher view on understanding of levels – autumn 2006
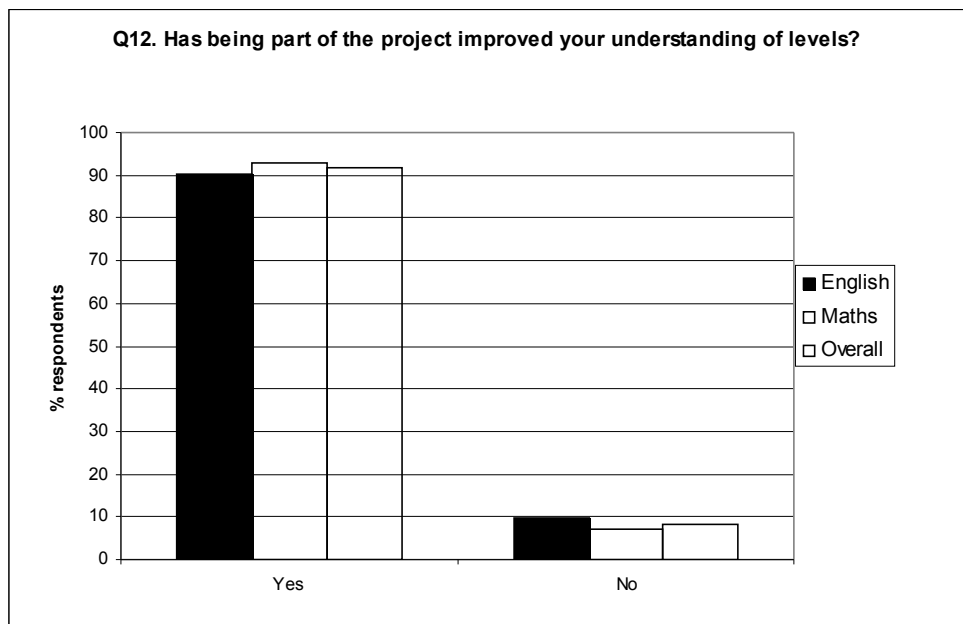
'Before APP, the teacher assessment levels I gave to my headteacher were based on little more than guesswork.'

'APP lets me get right under the skin of a level.'

'Before APP, I did not know about assessment focuses. Now I know how to move on to the next level.'

In the earliest questionnaire, teachers were asked to think about which of the processes that they had experienced as part of the project had contributed to their improved understanding. Chart 2.2 shows that it was the use of the guidelines with their own pupils which they had found most helpful,

followed by working with colleagues in school to agree and check their assessments through in-school moderation. It is interesting to note that this 'practice' was rated as far more helpful than the central training sessions.

Chart 2.2 What helps to improve understanding of levels – autumn 2006

Q12a. Which of these has been most helpful in improving your understanding of levels?
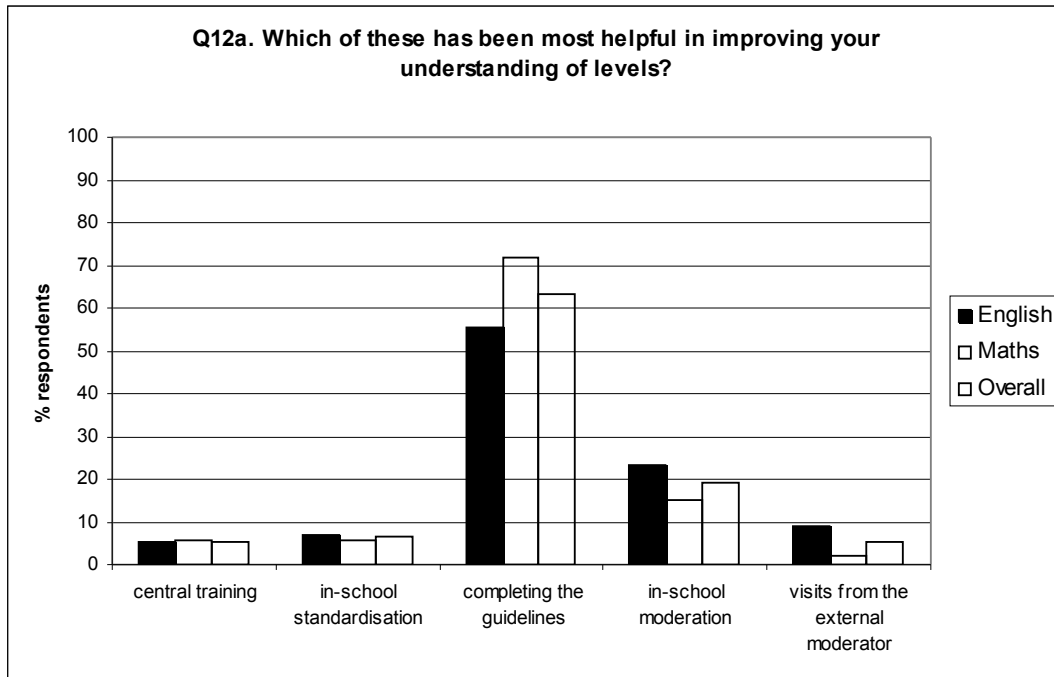

For some teachers, the use and application of assessment criteria across the full curriculum for mathematics, reading and writing presented some difficulties. When first introduced to the assessment focuses for the less familiar aspects of their subjects, misunderstandings and incorrect interpretations of the criteria were seen. For example, in the early moderation visits the English moderators reported occasional confusion between AF2 and AF3. The mathematics moderators also reported instances of teachers not understanding the guidelines. In some cases there appear to have been issues related to a lack of specialist subject knowledge or maybe a lack of confidence in trusting to professional experience in knowing what phrases such as 'in a range of' or 'with support' constituted.

There were also some specific issues with format and wording in the criteria. Negative statements within the criteria were considered unacceptable by teachers and, particularly for mathematics, some teachers struggled with differentiating between the APP level criteria themselves and the examples included of ways in which these criteria can be demonstrated by pupils. Materials were revised during the course of the pilot. The published assessment guidelines reproduced in appendix 2 have already incorporated these changes.
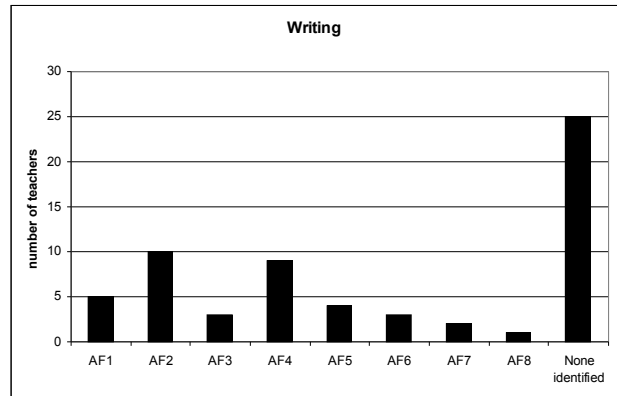
More challenging assessment focuses

In spring 2007, teachers were asked to identify the assessment focus for which they found it most difficult to make a judgement. Charts 2.3 to 2.8 show their responses.

Chart 2.3 Difficult AFs for reading

Reading
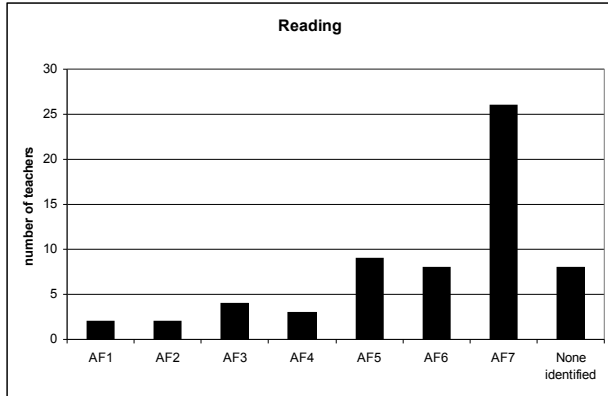
Chart 2.4 Difficult AFs for writing

Writing

Chart 2.5 Difficult AFs for Ma1

Ma1 Using and applying mathematics
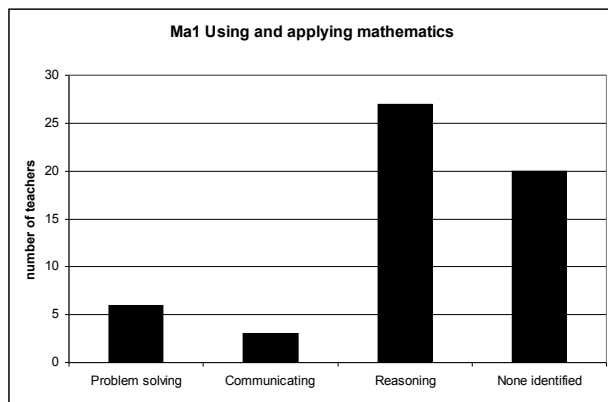
Chart 2.6 Difficult AFs for Ma2

Ma2 Number
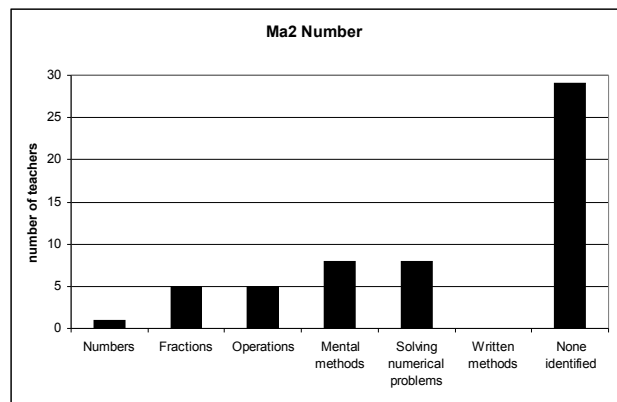
Chart 2.7 Difficult AFs for Ma3

Ma3 Shape, space and measures
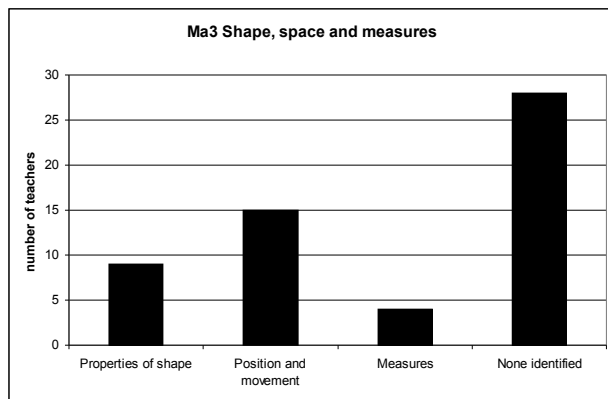
Chart 2.8 Difficult AFs for Ma4

Ma4 Handling data
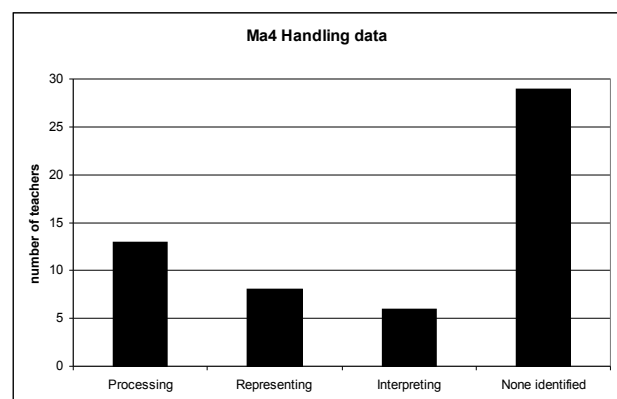
Teachers were asked through open-ended questions why these particular assessment focuses were problematic. Most referred to the difficulty in locating good quality evidence but these problem assessment focuses also correlate to the lack of understanding in some areas noticed by moderators.

Teachers initially focused on the process of completing the assessment guidelines and sometimes neglected to use either the guidance on making an overall level judgement or to refer to the

standards files when confirming their judgements. The use of the full range of APP materials was promoted by local authorities and through the visits from external moderators.

There was a perception that subject knowledge was improved by the use of APP, perhaps especially where teachers were not subject specialists in the chosen area to start with.

> 'Teacher knowledge has increased so therefore the children are being taught *all* the skills that they need to progress.'

> 'Teachers are more aware of the expectations for each level.'

Respondents have repeatedly commented on the positive impact that APP has on teachers' views of themselves as teachers, and both local authorities and teachers identified the value of APP in terms of professional development.

> 'Increased my confidence in teaching and I'm enjoying teaching more.'

The questionnaire for teachers in summer 2007 explored teacher views of impact on their teaching. More than 80 per cent thought that there had been changes to their teaching and, when asked for details, the most frequent response was to suggest that APP had resulted in more focused and informed teaching, backed up by better planning. A mathematics teacher commented that working at the level of detail demanded by APP assessments had a positive influence on his day-to-day teaching of mathematics.

Many teachers referred to the internalisation of the assessment focuses and how they are evidenced within the national curriculum levels as crucial to the success of APP. They felt that, once developed through practice, working with colleagues and taking part in moderation, this familiarity gave them an improved understanding of what achievement and progression should look like in mathematics, reading and writing.

3. The evidence base for APP assessments

Sound APP assessments rely on the careful review of relevant evidence. Several sources allow us to look at how teachers used evidence to support their APP assessments. These include teacher questionnaires, the analysis of assessment outcomes at different time points and the review of reports from moderators.

Initial teacher views on evidence

Feedback from teachers on their first assessment round suggested that finding and identifying evidence to feed into their judgements was a challenge. At early review meetings teachers referred to how difficult it was to look for evidence retrospectively and recognised that in subsequent rounds the process would be easier because they would be able to identify sources and opportunities as they occurred.

Chart 3.1 Making assessment judgements in first assessment round – autumn 2006
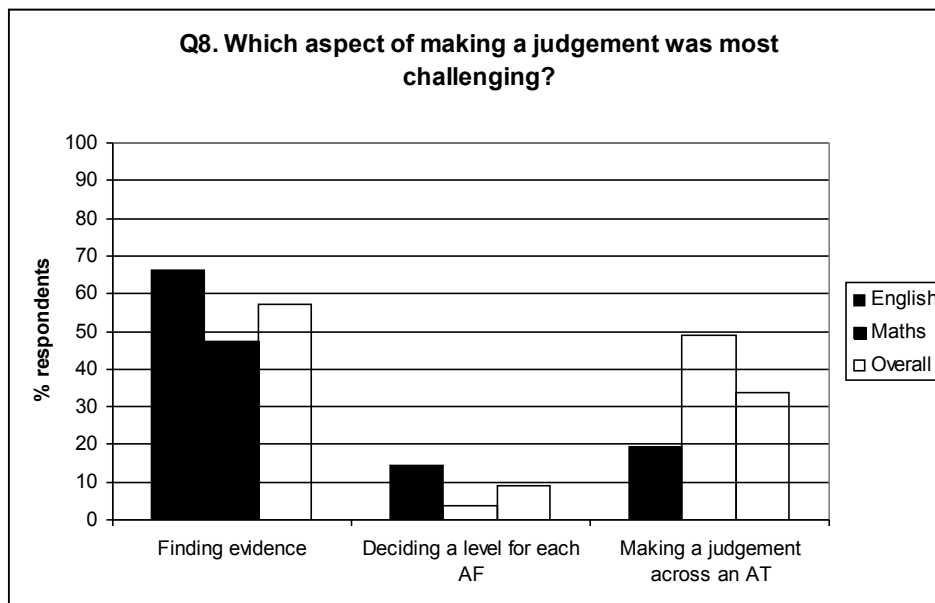
Moderators' view of evidence at the start of the project

In the first assessment round the focus of mathematics moderation was on Ma1 and Ma2. Moderators considered that there was usually just about sufficient evidence for the judgements in Ma2 that were shared with them but that overall evidence for Ma1 judgements was lacking. Teachers were said to be placing too much emphasis on the need for written evidence and showed reluctance to commit to a judgement where they lacked written evidence, even though when probed these teachers could give plenty of relevant information. Across the schools, moderators saw evidence from books, observations, notes from other subjects, marking, annotated plans, group work, discussions, starters and plenaries, children's own feedback and notes made by teaching assistants. Within individual schools, though, the range of evidence was often limited. Specific gaps in evidence existed for all Ma1 AFs and in Ma2 for 'Mental methods', 'Solving numerical problems',

and 'Written methods'. In general, there was little evidence of pupils having the opportunity to use choice or independence in their work.

In English, there was an almost universal report of insufficient evidence for reading where this had been the subject of moderation. In most cases there was a sufficient volume of evidence for writing although there were concerns about the range and degree of independence in the material presented. Most of the evidence for writing was biased towards narrative work, with just a few references to recounts, reports and instructions. There had been some attempts to assess work drawn from other areas of the curriculum and teachers were intending to develop this further. In reading, evidence was almost without exception presented as responses to questions; the lack of oral evidence was commented upon – one school had tried and given up as they found it too difficult. Specific gaps in evidence were noted: AF2, AF3 and AF4 in writing and AFs 4-7 in reading. Teachers had tried hard to access independent work to support their assessment. However, all moderators felt that the work they saw was too structured (sentence writing and punctuation exercises); it was often set as an assessed piece at the end of a teaching unit. In a number of cases the pieces of writing were too limited to provide a good basis for assessment, for instance a story opening presented out of context. There were also examples of work that had been produced with too much teacher support.

The assessment focuses which teachers had found to be the most difficult to assess (see pages 12 and 13) were often the ones where moderators were concerned about the evidence provided to support judgements.

'Insufficient evidence' judgements in APP assessment outcomes

Analysis of the APP outcomes from the first round of the data collection shows where gaps in evidence were so significant that teachers were unable to make judgements. Charts 3.2 to 3.7 were created by looking at the data across all the pupils in all year groups and for the full range of attainment targets.

Chart 3.2 Distribution of Ma1 AF levels for whole sample – autumn 2006
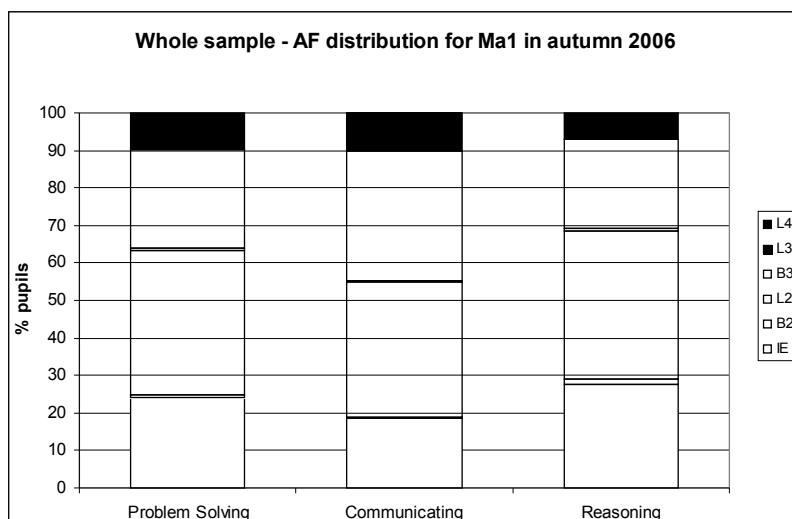
The Ma1 assessment focuses were consistently identified by teachers as being difficult to assess and the chart shows that 'insufficient evidence' was recorded as the assessment outcome for reasoning for just under a third of pupils when assessments were made at the end of the first autumn term of the pilot.

Chart 3.3 Distribution of Ma2 AF levels for whole sample – autumn 2006
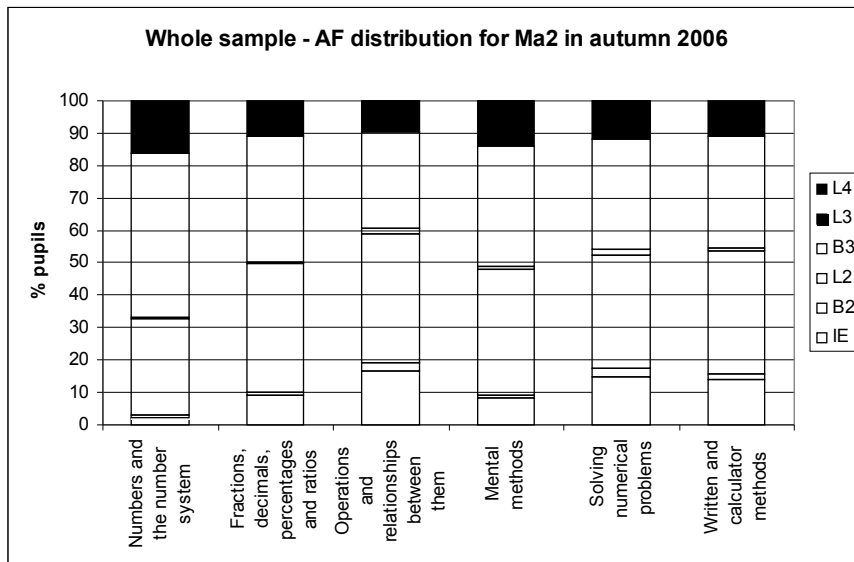
A lack of evidence was less of a problem across the assessment focuses for Ma2, which is recognised as the most significant contributor to overall attainment in mathematics in statutory QCA guidance on assessment. However, it is worth noting that the two assessment focuses which require the application of understanding, 'Operations and relationships between them' and 'Solving numerical problems', were the most likely to be recorded as having 'insufficient evidence'.

Chart 3.4 Distribution of Ma3 AF levels for whole sample – autumn 2006
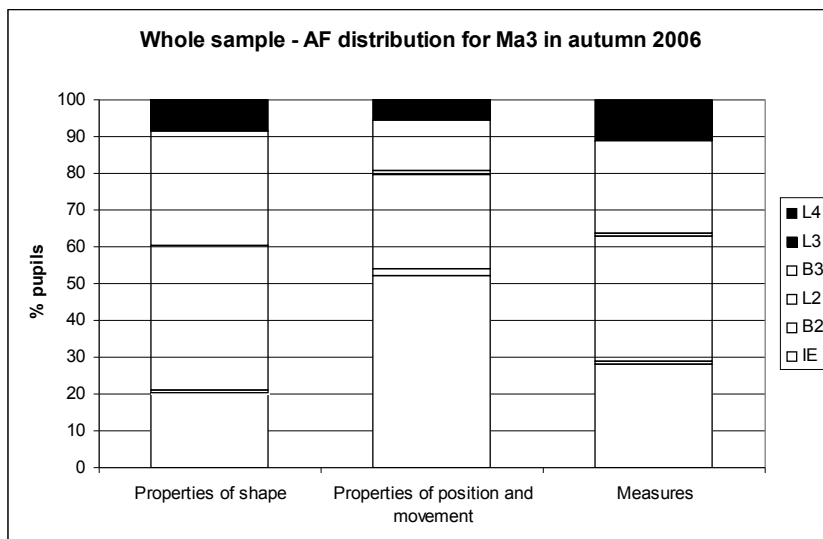
Chart 3.5 Distribution of Ma4 AF levels for whole sample – autumn 2006
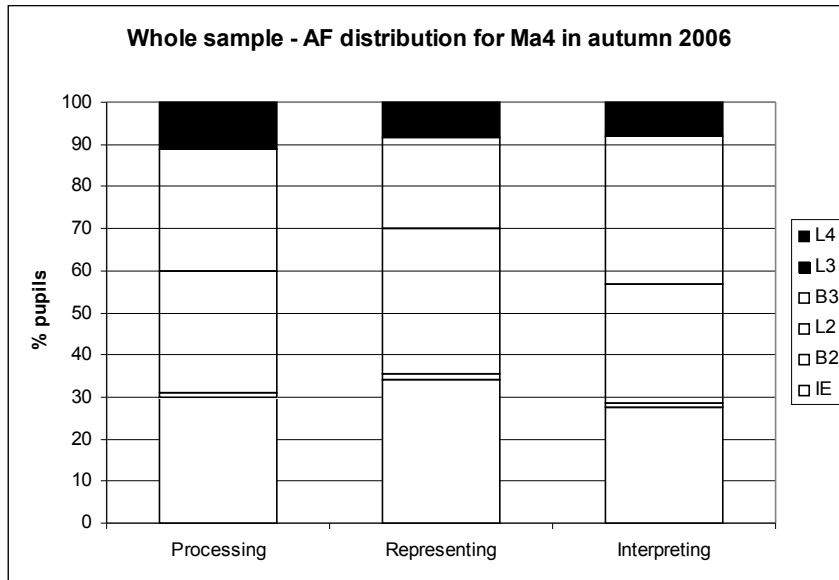
Teacher comments suggested that there were problems with finding evidence for Ma3 and Ma4 because 'it was the wrong time of year', as these aspects of mathematics were not covered in plans for the autumn term.

In Ma4, 'interpreting' was generally the assessment focus where there was more likely to be evidence available. During a moderation visit where this was discussed with a teacher, she gave a candid explanation that since 'processing' and 'representation' did not 'come up' in the national curriculum tests at the end of key stage 2, they did not cover these with pupils.

Chart 3.6 Distribution of reading AF levels for whole sample – autumn 2006
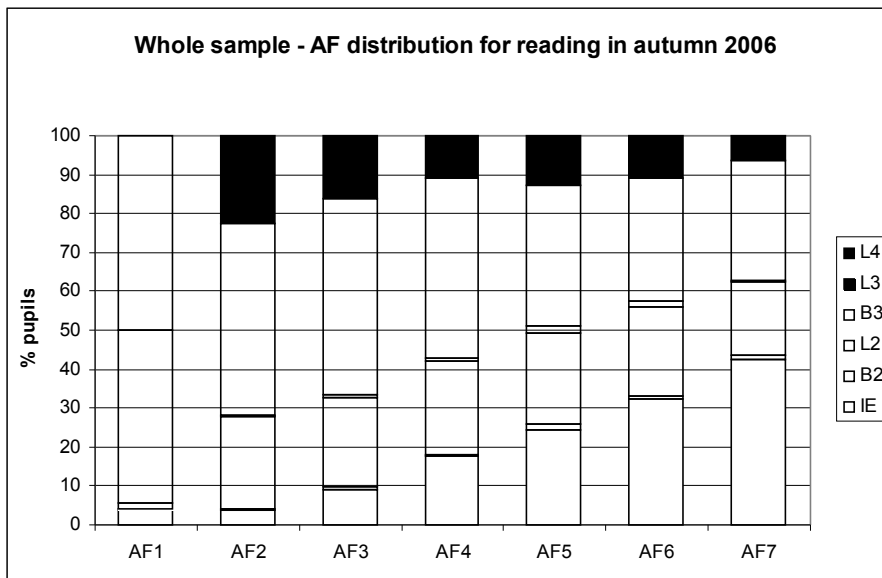
As teachers identified in their questionnaire responses, the assessment focuses associated with higher order reading skills were the most difficult to evidence. More than 40 per cent of APP assessments had insufficient evidence recorded for AF7 – 'relate texts to their social, cultural and historical contexts and literary traditions' – and more than 39 per cent had insufficient evidence for AF6 – 'identify and comment on writers' purposes and viewpoints and the overall effect of the text on the reader'.

Chart 3.7 Distribution of writing AF levels for whole sample – autumn 2006
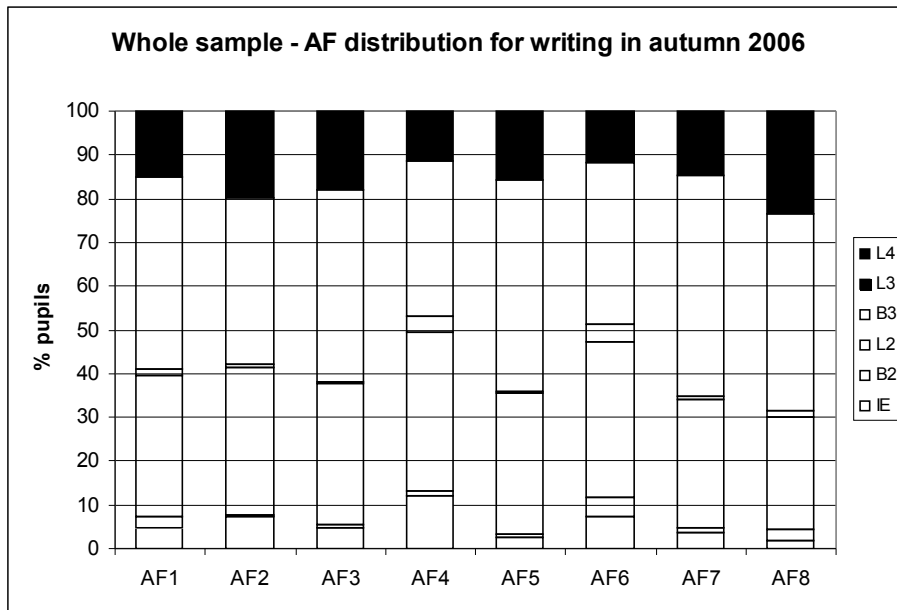
For writing, teachers were much less likely to record 'insufficient evidence' as an outcome but when they did, AF4 – 'construct paragraphs and use cohesion within and between paragraphs' – appeared to be the most problematic. This may have been related to the lack of extended writing opportunities referred to in moderator reports.

Seeing these gaps in the evidence base for their assessments was very revealing for teachers. In questionnaire responses they indicated that this was the chief lesson learned from this early stage in the project.

Chart 3.8 Learning from the first round of APP assessments autumn 2006

Q13a. Learning about the class as a whole

Category	Value
Uncovered gaps in curriculum/teaching coverage	~41
Improved understanding of next steps/progression	~21
Became more aware of AFs and their contribution to NC levels	~17
Improved my questioning with all pupils	~13
Weaknesses in the focus group symptomatic of weakness in class as a whole	~11
Learned about importance of independent work to show where pupils are	~9
More understanding that pupils have different strengths and weaknesses, even when working at the same level	~3

Many teachers provided specific details of where their identified gaps were:

> 'I feel my teaching has had too much emphasis on Ma2. I need to do more on Ma1, 3 and 4.'

> 'It's highlighted assessment focuses in reading that have been neglected.'

Changes to the evidence base after using APP

After the second round of assessments, both English and mathematics moderators reported a general improvement in the quality and range of evidence, with encouraging references to the use of evidence from other subject areas. There were some examples of creative and innovative use of evidence. In mathematics, there were instances of work on shape carried out in physical education (PE) lessons, problem-solving activities as part of a joint venture with a local secondary school and investigative work involving parents. Concern was still evident about the lack of evidence for Ma1, the limited quality of evidence provided for reading and the extent to which pupil work in both English and mathematics could be considered to be independent.

Mathematics moderators reported that teachers were becoming more likely to recognise the importance of oral evidence but that in general teachers were still likely to err on the side of caution in awarding levels because they did not attach sufficient value to evidence which was not 'written down' by the pupils.

For a number of assessment focuses, very useful evidence from other curriculum areas can be used to add to teachers' understanding of what pupils can do, particularly since, away from subject focus lessons, there is often more opportunity for independence and choice. This idea was promoted to teachers at initial training and by the moderators, and in spring 2007, teachers were asked to indicate if they had been able to draw on such evidence and to give details.

Chart 3.9 Evidence from other subjects – spring 2007

Q14 Have you drawn on evidence from other subjects?
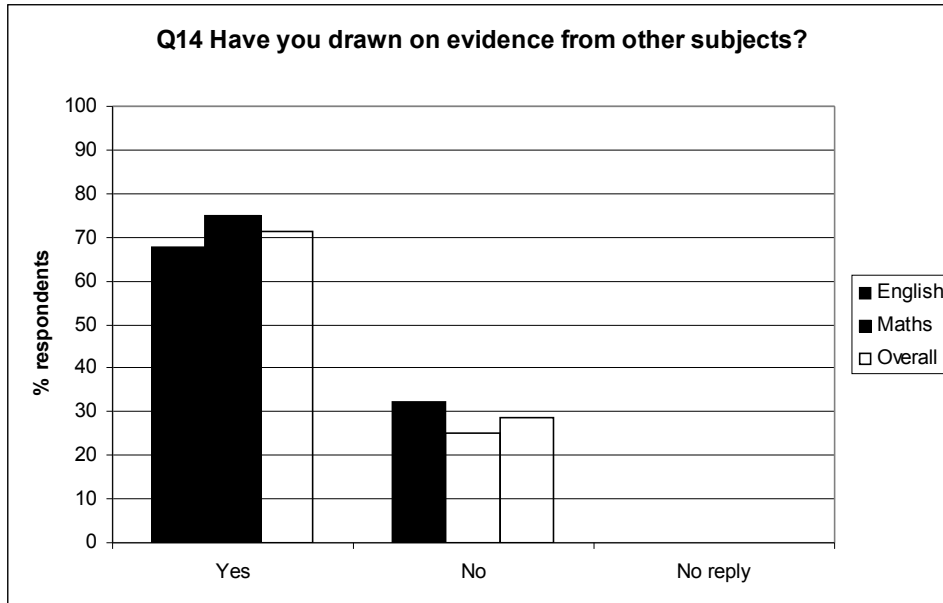
Mathematics teachers had used evidence from information and communication technology (ICT), geography, history (for example, directions linked to a project on mazes), design and technology, personal, social and health education (PSHE), art, religious education (RE) (for example percentages during a project on poverty in the third world), PE, English and, most frequently, science. English teachers had used evidence from ICT, RE, geography (for example retrieving information from a project on the Gambia), history, design and technology (for example report planning), PSHE and art. From moderator reports it appears that there is considerable school-level variation in the approach to the use of evidence from other subject areas. In particular, middle schools with secondary-style timetabling and other schools where teaching takes place in sets tend to find it difficult to pick up evidence from other subject areas and some make no attempt to do so.

The improvement in the amount of evidence provided to support judgements reported by moderators was reflected in the analysis of the APP assessment outcomes in the project data collection for the second and third terms' assessment focus level distributions. A full set of charts showing assessment focus level distributions in autumn and summer assessment rounds from the first and second years of the pilot is provided in appendisx 3. Charts 3.10 and 3.11 show assessment focus distributions for reading and Ma1, from the assessments undertaken in summer 2007. They illustrate the notable reduction in the number of pupils for whom 'insufficient evidence' was recorded in the final round of the first year of the pilot compared to the first (see pages 16 and 18).

Chart 3.10 Distribution of reading AF levels for whole sample – summer 2007

Whole sample - AF distribution for reading in assessment round 3
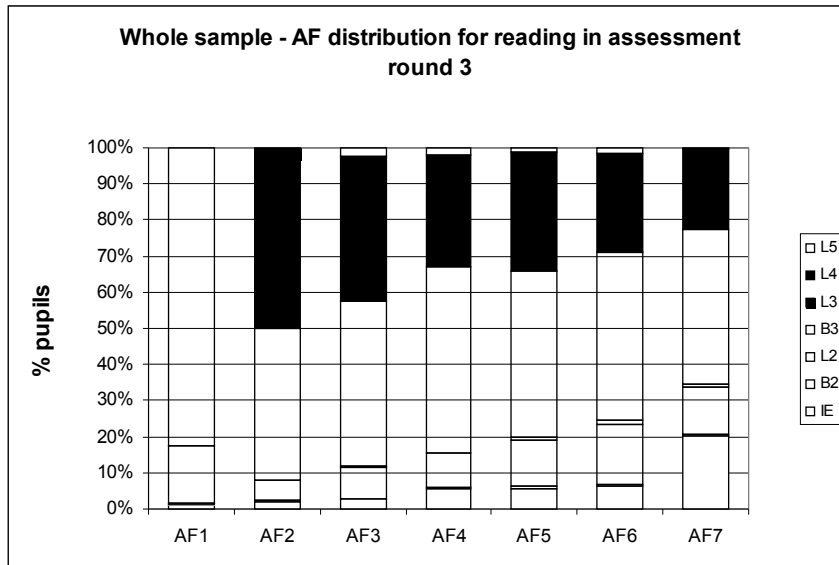
Chart 3.11 Distribution of Ma1 AF levels for whole sample – summer 2007

Whole sample - AF distribution for Ma1 in assessment round 3
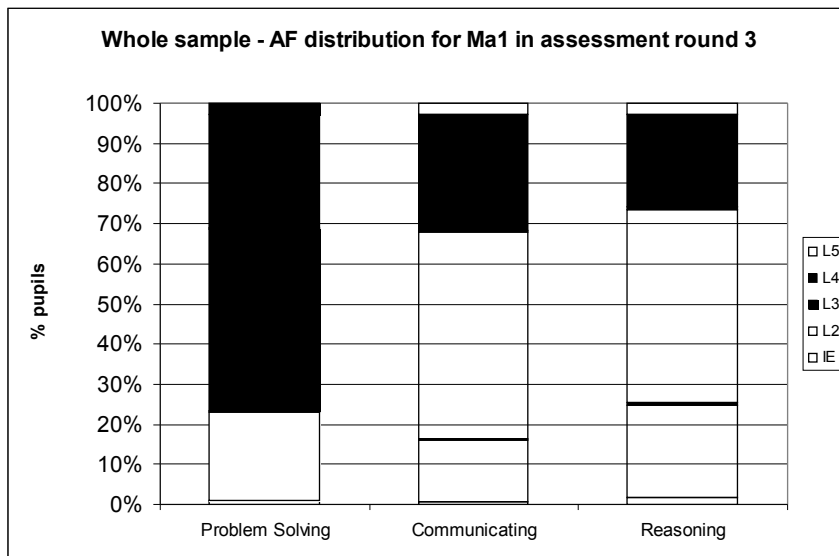
By the summer term, the proportion of pupils for whom judgements of 'insufficient evidence' were recorded had reduced to less than 5 per cent in mathematics and, with the exception of AF7 (which had been reduced to 20 per cent from 42 per cent), to less than 10 per cent for the reading assessment focuses.

From the moderator perspective, the third term visits revealed some improvement in the extent to which the assessment evidence included work demonstrating independence and choice. For writing, there were more examples of work where pupils had been given some freedom in creating ideas or selecting the way in which their ideas could be presented. There were still only occasional examples where a choice of form or purpose was offered to the writer. Fewer examples of heavily

scaffolded work were seen and there were also more complete pieces of writing. The difficulties in locating evidence to support reading assessment were still apparent; throughout the year moderators had reported over-reliance on closed-style comprehension tasks. In summer there was more evidence of open-ended activities and some improvement to questioning styles. However, in a number of schools, teachers included responses from the optional tests in their samples of work and there was a feeling, articulated at some of the moderation visits, that this was all that was available in the summer term because of the pressure of time and the focus on testing.

Generally, the mathematics moderators considered that there had been an improvement in the extent to which pupils were given the opportunity to make decisions in their mathematics and to work independently, but that this varied noticeably from school to school. Some teachers interpreted 'independent' as 'without adult support' and did not recognise that, for effective assessment of a skill to take place, there must be some distance between the teaching input and its application by the pupil. Similarly, offering more choice to pupils, even in small ways, can provide sound opportunities, but only a minority of teachers recognise this. In many cases, though, teachers took the opportunities for assessment offered by geography or science, where pupils could choose how to apply their mathematics.

> 'Differing views of what 'choice' looks like, from letting children decide how to perform a calculation, right through to asking them to devise an activity, decide how to tackle and evaluate it and then how to present and record their findings.' (moderator).

Evidence to support assessment in the second year of the pilot

In the second year of the pilot, when almost the whole cohort of pupils in the data collection sample was replaced and teachers new to APP were contributing both to the data collection and to moderation activities, the story was similar to the first year. Over the year there was a developing awareness of the nature and range of evidence required. Appendix 4 includes assessment focus level distribution charts from the beginning and end of the year and there was the same shift away from the recording of 'insufficient evidence' as was seen in the first year of the pilot.

Changes to the evidence base for judgements were also apparent in the reports prepared for schools after each of the moderator events. In their reports moderators give ratings for the evidence provided. Charts 3.12 to 3.14 compare the moderator views of evidence presented by schools after one, two or three rounds of moderation activity and show that, with increasing experience of APP assessment, judgements were supported by more evidence, drawn from a wider range of sources and included more examples of work where pupils showed independence and choice.

Chart 3.12 Moderator views on the amount of evidence to support APP judgements

Moderator feedback in second year of pilot - sufficient evidence to support judgements in...
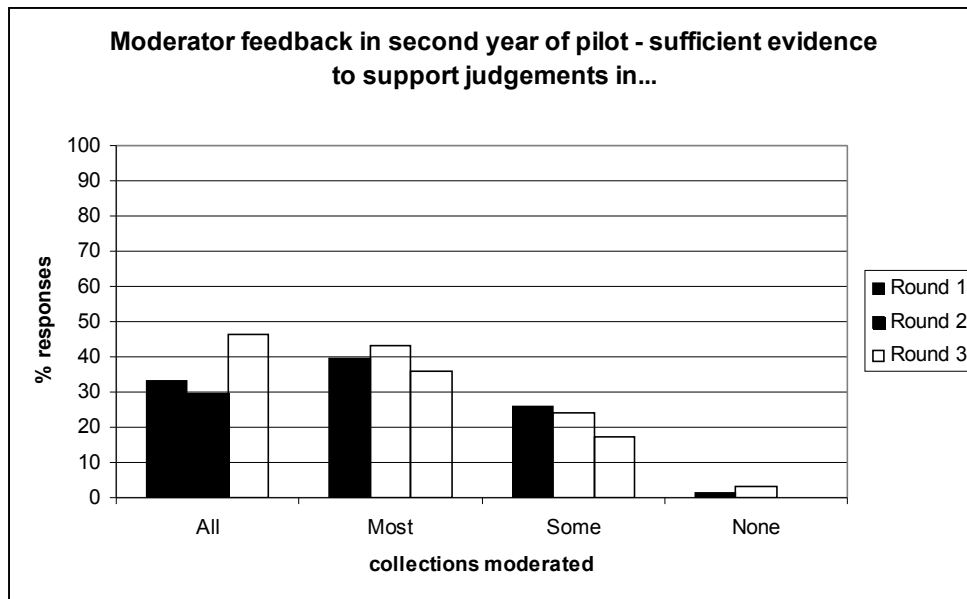
The greatest improvement came in the third round of moderation, when all participating schools were looking at attainment targets which they had already worked on. (It is important to remember that the data from the moderation events includes non-pilot schools as well as those in the APP pilot.)

Chart 3.13 Moderator views on the inclusion of 'independent' work in evidence

Moderator feedback in second year of pilot - evidence presented included...
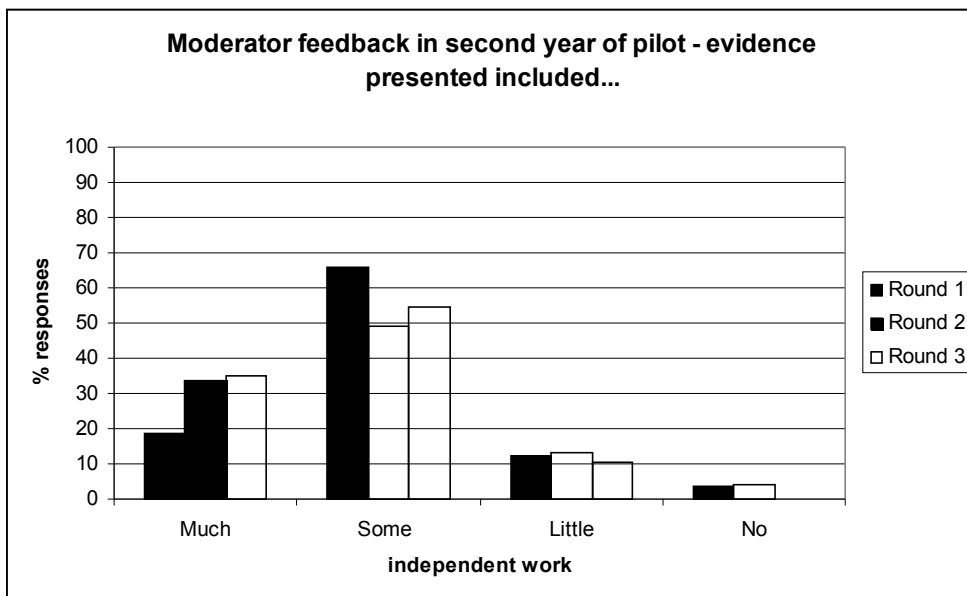
Chart 3.14 Moderator views on the range in evidence presented to support APP judgements

Moderator feedback in second year of pilot - range of evidence presented was...

Scale 1 = very wide to 4 = limited
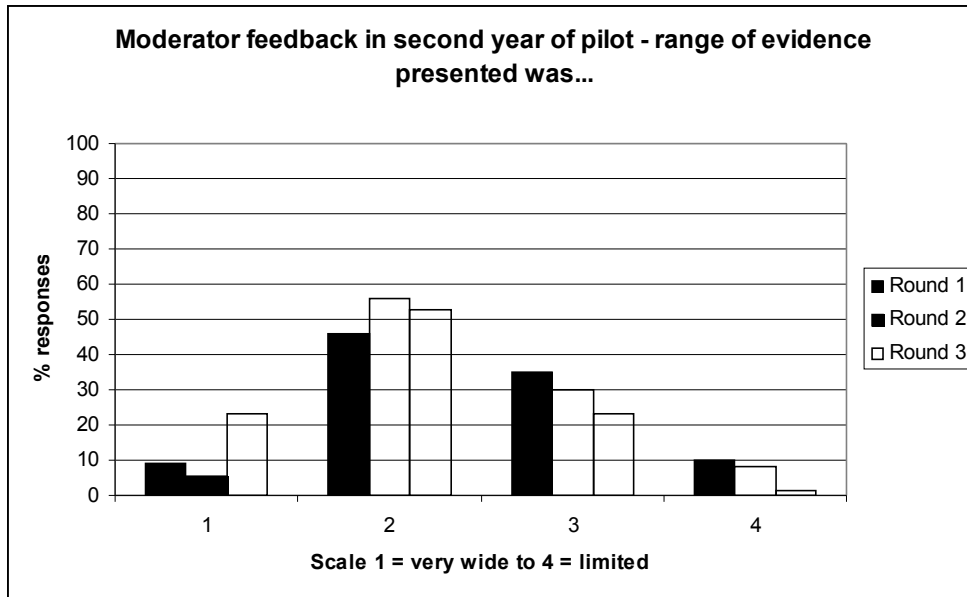
It is apparent then that carrying out APP assessment prompted teachers to take action which meant that subsequent assessments could draw on a wider and improved evidence base. Faced with the obvious message that they lacked evidence for particular assessment focuses there are two major implications:

• there is no evidence because existing assessment activities do not address these assessment focuses

• there is no evidence because the curriculum offered is not sufficiently broad and does not deliver experiences which allow to pupils to develop the skills which the assessment focuses are designed to highlight.

A third explanation seemed to be relevant when assessing mathematics, that there was sometimes no evidence because the part of the curriculum which these assessment focuses relate to was covered cyclically and assessment took place at an inappropriate time. Teachers comments along the lines of 'we don't do shape until the summer term' were very frequent. Of course, outside the pilot, the timing of assessments should be driven by the needs of pupils and teachers for the knowledge to inform next steps in learning and would not be constrained by the requirement to return assessment data by a specific deadline.

The next two sections explore the actions taken by teachers which may have contributed to the improvement seen in the evidence base.

4. Developing skills in everyday assessment activities

Through observing meetings for teachers throughout the project, it has been clear that their ideas of what constitutes a 'good' assessment activity have changed considerably. In the second questionnaire to teachers, there was an overwhelming view that they were using a wider range of approaches to assessment as a result of their experience of APP.

Chart 4.1 Teacher response after second round of APP assessment

Q12d Using MCP has prompted me to widen the range of activities I use for assessment
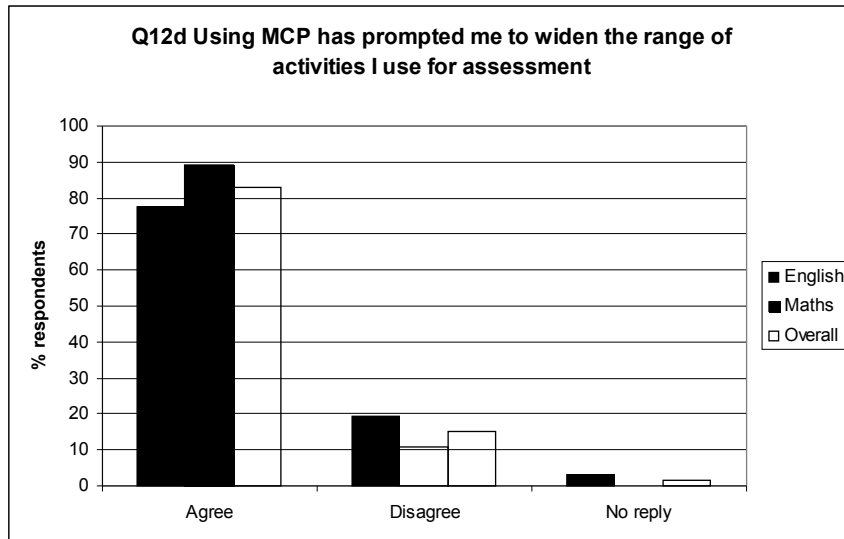
Prior to using APP, there was a reliance on testing in a range of forms to measure achievement and progress. At the initial training meetings, the message that the more ephemeral evidence of what children can do, which is often what teachers simply 'know' about their children, can be just as valuable and sometimes more informative than written test results, was welcomed with enthusiasm. Using the APP guidelines, however, is not simply a matter of ticking a box because that's 'what I think'. The criteria exist to promote rigour and consistency in assessment and applying this structured analysis does mean that teachers have to be able to check how it is that they know what a pupil can do or understand in relation to a specific assessment focus. So when making first assessment judgements as seen in section 3, teachers often struggled to identify where their evidence came from. Some teachers have continued to struggle with the challenge this presented to them throughout the project. Interestingly, the idea that observation of pupils in everyday classroom situations can yield a wealth of evidence was far easier for the teachers in the key stage 1 APP pilot to take on than it was for their colleagues in key stage 2. Perhaps this was because the APP has much in common with the recommended approach for teacher assessment at the end of key stage 1 and perhaps because there is less embedded use of testing within and at the end of each year in key stage 1.

From the first questionnaire, teacher comments showed that they were starting to appreciate that a change of approach was needed in their use of assessment activities to help them get at the evidence needed.

'The questions I was asking tended to limit the children's answers.'

'It opened up my idea of assessing for reading and broadened my ideas for how to approach guided reading.'

'I need to look at cross-curricular activities to provide evidence for Ma1.'

Many teachers appreciated that conversations with pupils could provide really good insights into understanding but sometimes when these opportunities were created, they tended to use a set of planned questions to support the conversation rather than letting it develop more naturally. This often limited the value of what they could find out. Over time there were indications, particularly through the evidence shared at moderations and through teacher comments, that with greater familiarity with the assessment focuses they could 'tune in' and make the most of incidental opportunities, rather than having to 'construct' what in the most extreme cases amounted to an oral 'test' for a target assessment focus.

'We're just open to picking things up now; don't always have to get hung up on creating assessment opportunities.'

'I was writing questions to ask the children and I was only giving them the opportunity to give me a level two answer and just by re-jigging the questions, you can actually give them an opportunity to give much more sophisticated answers and you can use more of the assessment focuses by giving them a more broad question.'

The last comment shows that, with increasing familiarity and confidence, teachers also realised that an activity or piece of work could often provide good evidence for more than one assessment focus. This meant that the assessment activity was probably more meaningful and coherent and also that 'evidence gathering' was more efficient.

Chart 4.2 Teacher view on planning for assessment – spring 2007
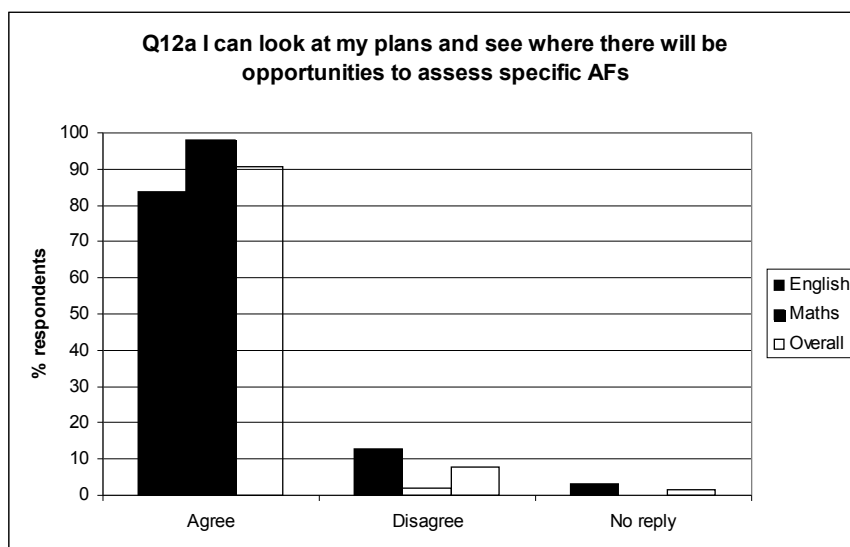
By the second term of the project, teachers were using their planning to look forward, so that they could see where the naturally occurring opportunities would be to pick up useful information for APP assessment.

'I matched each lesson with the assessment guidelines.'

'Greater focus on guided reading with more detailed planning of questions before the lesson.'

'I added some small group assessment outside numeracy lessons.'

'Planned more opportunities for children to reason about maths.'

There was also a recognition that you can sometimes learn more about how to move children on by knowing what they can't do, and in order to find that out you have to build an element of challenge into the activities which contribute to your assessments.

'Maths has been under a cloud of right or wrong till now.'

'Children who began the year with negative and defeatist attitudes have been encouraged to tackle challenges and accept that initial failure can be necessary and valuable.' (mathematics teacher)

Of course, not all teachers have moved forward in their practice at the same rate, but the process in the second year where teachers met for group moderation has often been very helpful in promoting change. It is far more effective for teachers to learn through sharing ideas with other colleagues who have already changed their practice, than to hear messages from external moderators, however valid these are. The contrast between the type of activities offered to pupils working at the same level in different schools was often very striking and it was possible to see teachers actually taken aback when they realised that they might be limiting what their pupils could do by the type of assessments used.

In the questionnaire responses from teachers at the end of the project, 92 per cent of respondents indicated that they believed APP had significantly improved their ability to identify good assessment opportunities in everyday classroom activities.

How these insights are recorded has significance for the manageability of APP; this issue is returned to in section 8 of this report.

5. Making changes to classroom provision

Within the project, teachers were asked to carry out APP assessments for a small focus group of pupils, but the intention was that what they learned from the process would inform teaching and learning in their class as a whole. After the first assessment round, teachers were asked whether they had picked up useful information about their class as a whole and to give some detail on what they thought they had learned.

Chart 5.1 Information gained from APP – autumn 2006
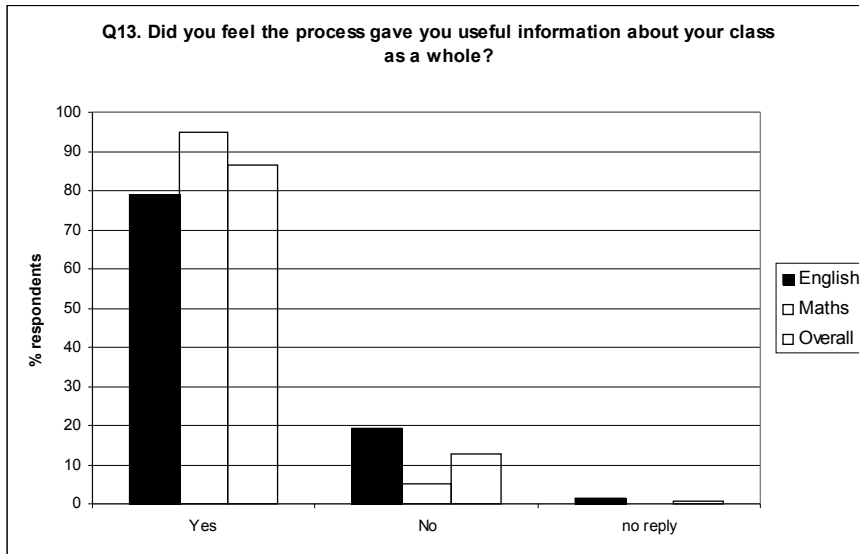
In most cases this information related to gaps in teaching and learning.

Chart 5.2 Nature of information gained from APP – autumn 2006
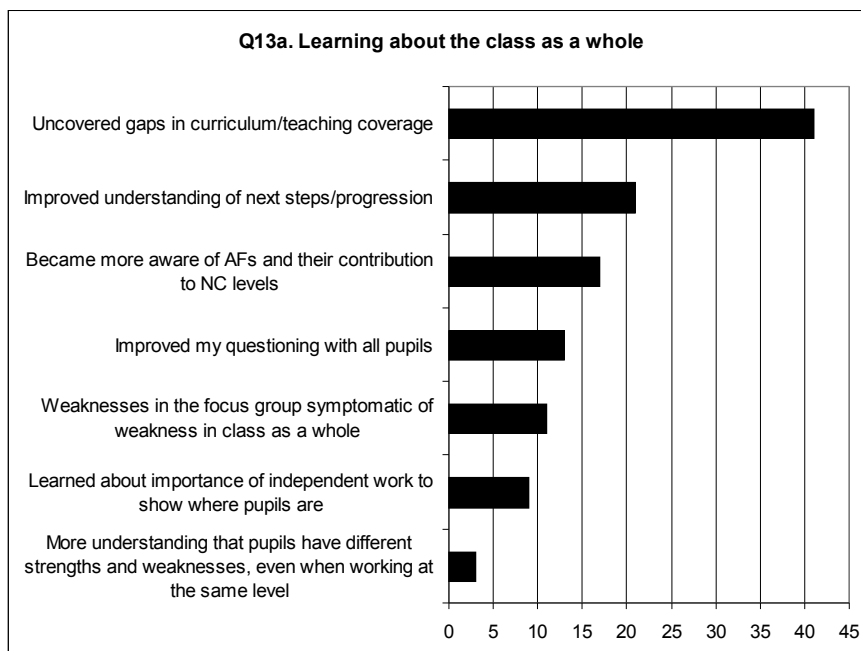
Teachers' individual comments added detail:

'Not all children are having the chance to demonstrate their knowledge and understanding.'

'It highlighted weaknesses in the learning which might reflect weaknesses in the teaching of English.'

'It changed my view of low achievers and assessment focuses other than AF1 and AF2.' (reading)

'Some areas need to be taught in different ways, as children are not remembering them at a later stage.'

A follow-up question asked if they planned any changes to their teaching as a result of what they had found and their responses were grouped and shown on the chart below.

Chart 5.3 Intentions for changes to planning – after first APP assessment
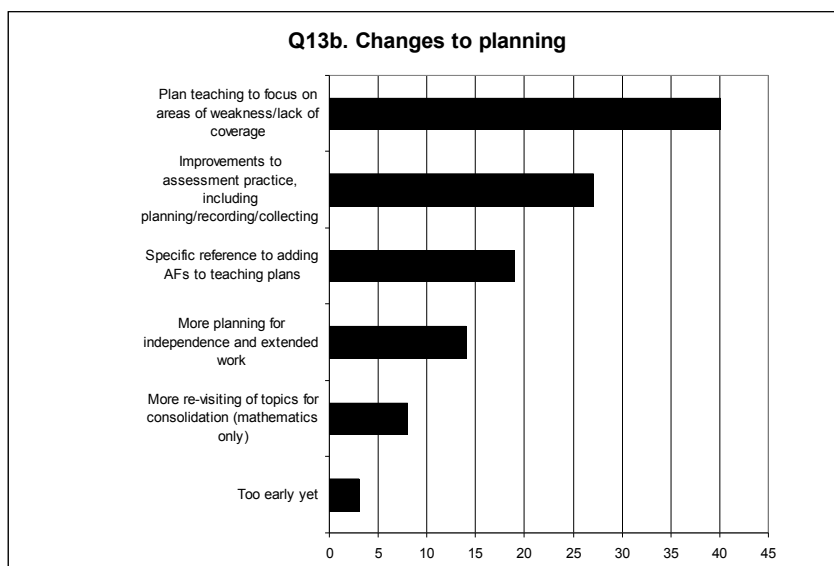
Teacher comments:

'Yes, lots – more work on angles, shape, space and interpreting data which is an area for development.'

'Yes, plan extended problem solving activities and more data handling work.'

'More narrative writing is needed.'

'To record more in children's books as evidence.'

'Change guided reading in particular to address all the assessment focuses.'

'I should incorporate more work on verb tenses.'

'Yes, adapt planning to address areas which don't normally appear in year 6 scheme of work but the children need to master to move on.' (mathematics)

'Have introduced problem solving books.'

'Changing groupings.'

After a second round of assessment, teachers were asked if they had gone on to make the changes to their teaching plans.

Chart 5.4 Implemented changes to planning – spring 2007

Q11. Did you adapt your teaching plans after the last assessment?

Q11 Give brief details of changes to your teaching plans
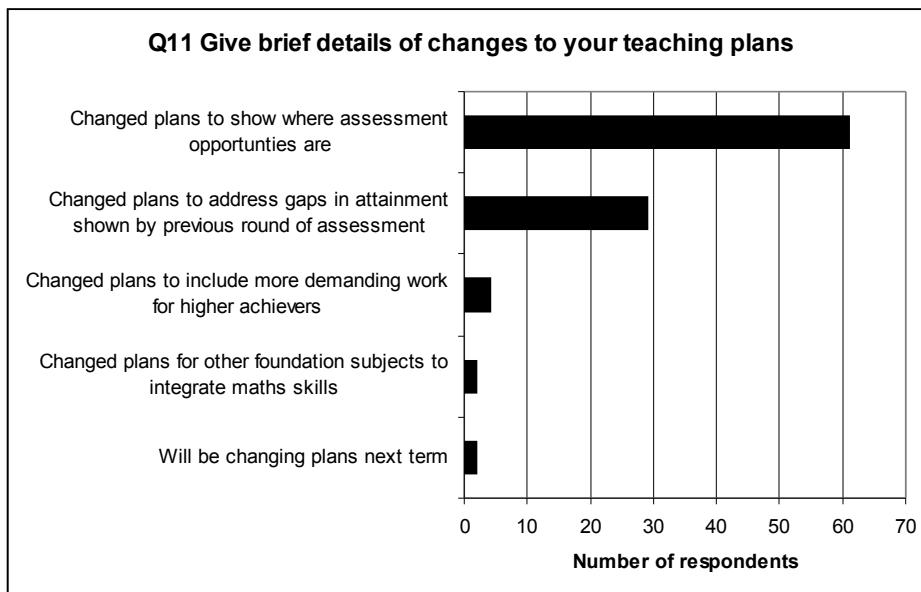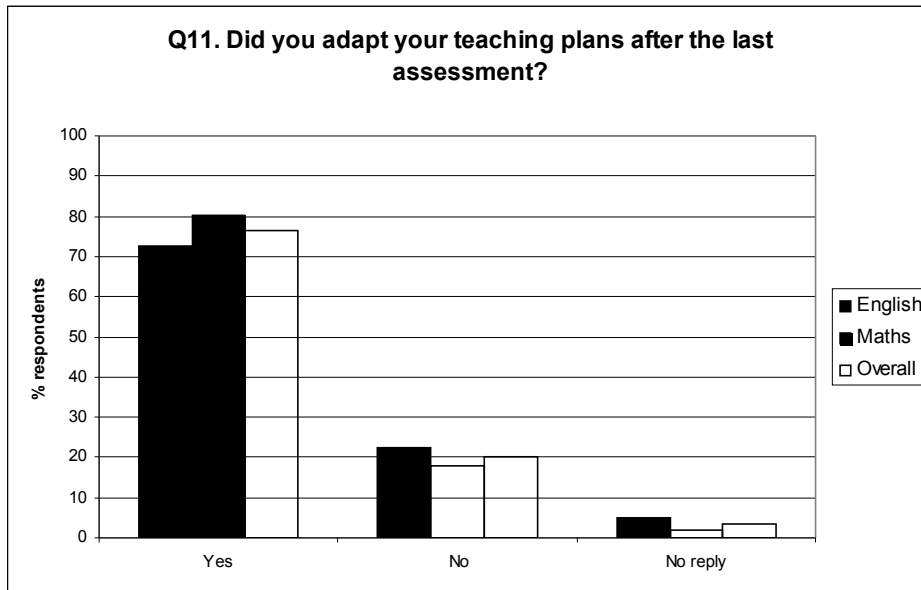
While most teachers seemed to be concentrating on planning for assessment opportunities rather than addressing any shortcomings in provision, there were some comments which showed awareness of what needed to be done to improve learning experiences.

'Used more focused comprehension work to try to raise inference skills.'

'Planning included more work on connectives and linking paragraphs and vocabulary as these were weaknesses identified in last round.'

'Planned to cover areas in more detail that had only been touched on and looked to next level to extend more able.'

'Looked within other foundation subjects to integrate and promote maths skills.'

In the summer term 2007, teachers were asked again what changes they had made on the basis of assessment outcomes from the previous round. Most responses referred to attempts to home in on identified areas of weakness, for example inference and deduction, paragraphing, fractions and decimals, and investigative work. The next most frequent comment related to attempts to improve coverage of previously neglected areas:

> 'We're not doing enough explicit teaching of reading at key stage 2.'

Teachers in mathematics were more likely to refer to covering these neglected areas and their focuses included 'Problem solving', 'Measures', 'Percentages' and 'Data handling'. There were a few references to improving differentiation in mathematics, especially for more able pupils.

Feedback from the final questionnaire

In the final questionnaire, teachers were presented with some aspects of teaching and assessment which might have been affected by their experience in the pilot and asked to say whether or not they considered that these had been improved by APP. Three of these aspects were related to the broadening of curricular experience. Chart 5.5 shows the teacher response. In this chart the category 'not relevant' was used where APP was only being used for one of the subjects.

Chart 5.5 Teacher views on impact of APP – summer 2008
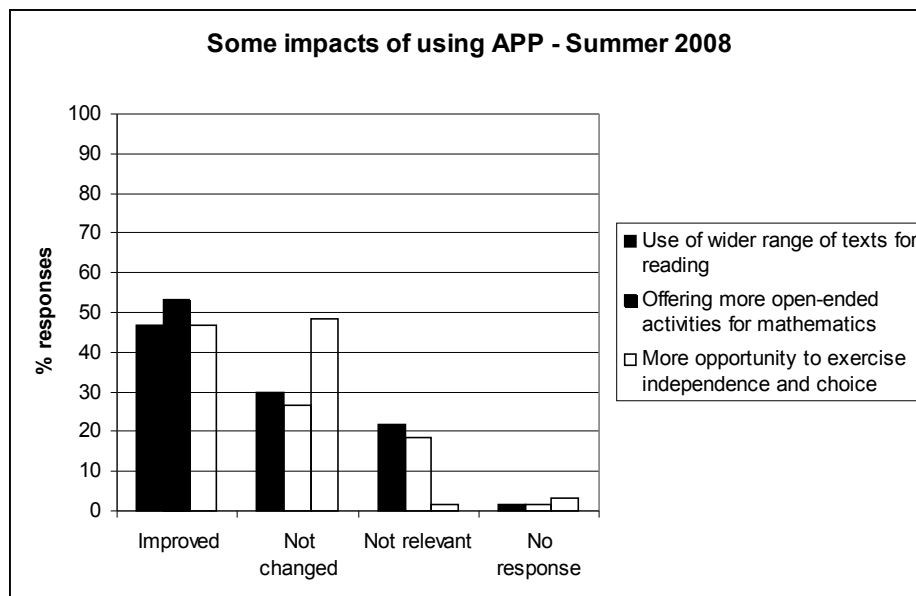
Teachers' views were quite divided on the extent to which these aspects of practice had been improved by APP, compared to aspects which have been covered in other sections of this report, such as their ability to identify gaps in learning or to understand the characteristics of performance at different national curriculum levels. These three aspects all require real change to classroom provision and it may be that there is a gap between identifying a need and being able to address it. Additional scrutiny of the negative responses did not reveal any relationship to the level of support available in the local authority (which might have contributed ideas about how to move practice forward) or to the number of teachers in the schools using APP (working in the pilot in a more

isolated way might limit the ability to initiate change). However, these negative responses were more likely to come from schools which also said that they had not replaced existing assessments with APP. This could indicate that there is less willingness to change practice where APP is seen only as an adjunct to test-based assessment or it may be the case that some schools simply have a greater 'openness' to change which has allowed them to move forward more rapidly with APP.

Relationship to teachers' planning

By summer 2007, the way in which APP could help in the integration of the cycle of 'plan, teach, assess, review, plan' was becoming more obvious to many of the teachers.

> 'It's changing the focus from teaching for assessment to assessment for teaching.'

Headteacher responses to a questionnaire in the summer term showed that 95 per cent considered that APP assessment outcomes had been used to inform planning and 60 per cent of headteachers thought that pupils were now experiencing a greater range of learning opportunities.

Recognition of how well APP supports planning for effective teaching was a key feature of the project throughout. In the final questionnaire, teachers identified that this was the most easily managed aspect of the whole APP assessment process. In summer 2007, some teachers expressed concern that APP would not sit well with the new frameworks for mathematics and English introduced from autumn 2007. This issue was explored in the final questionnaire in summer 2008, by which time 89 per cent of the schools responding had adopted the new frameworks and almost all of them said that they worked well together.

> 'Clear link with objectives – seamless.'
>
> 'Helps to make sure we assess reading and writing in every unit.'
>
> 'Plan using frameworks, assess against APP then re-plan using outcomes.'

6. Attainment and progress

The section draws on the analysis of assessment outcome data collected throughout the two years of the project. It also explores the perceptions of teachers and headteachers of any possible impact of APP on attainment and progress.

Levels achieved in reading, writing and mathematics as measured by APP

The analysis of assessment outcomes throughout the two-year pilot has provided a complete set of data on the patterns of attainment within key stage 2. Appendices 5 and 6 include all the charts for this data. To be useful from a school perspective, level judgements made from APP assessments would need to support tracking and monitoring of pupils' progress over time. In this section, emerging progress data is explored to see whether progress, as measured by APP assessments, gives a sensible and defensible picture of the way pupils progress within key stage 2.

At each assessment round, teachers derive an overall level for reading, writing and the four attainment targets in mathematics based on individual assessment focus judgements. Within the overall level, the teachers are then asked to refine their judgement by deciding whether the pupil's attainment is best described as high, secure or low within their level using the guidance on the flowchart. We know that teachers are often required to provide teacher assessment judgements in terms of 'sub-level'. Refining APP judgements within levels supports teachers and schools in tracking progress within a level. It has often been the case that teachers have equated the within-level APP judgement and the more familiar sub-levels derived from tests.

For mathematics, the data collection agency applied a formula based on different weightings for the individual attainment targets (as set out in QCA advice on assessing mathematics) to calculate the overall level for the subject.

The charts show the distribution of levels in appendices 5 and 6 for each year group; the change over a year shows pupil attainment moving into the higher levels following a normal distribution pattern. It was also important to look at whether APP assessments for individual pupils show consistent progress as this gives some insight into the reliability of APP assessments.

Tables 6.1 and 6.2 show a summary of changes in overall level for all pupils for whom measures were available between assessment rounds one and three in the first and second years of the pilot.

It would not be appropriate to place too much emphasis on the exact measurements of progress within this analysis, as it is important to remember that, in the context of the pilot, the accuracy of teacher judgements has changed over time as they have become more familiar with the assessment process and have been able to draw on better evidence. However, using the level judgement outcomes from APP does appear to provide a source of data that could be interrogated to reveal patterns of progress, provided that such outcomes are available for all relevant pupils.

Table 6.1 Summary of APP progress analysis between assessment rounds – 2006/7

	Year 3						Year 4					
	Jan–Mar			Mar–May			Jan–Mar			Mar–May		
Per cent	same	up	down	same	up	down	same	up	down	same	up	down
Reading	45	51	3	25	68	6	36	57	7	30	64	6
Writing	35	60	5	38	56	6	35	57	8	33	62	5
Mathematics	29	71	0	25	73	2	31	68	2	36	63	1

	Year 5					
	Jan–Mar			Mar–May		
Per cent	same	up	down	same	up	down
Reading	31	58	9	25	68	6
Writing	34	56	10	42	52	6
Mathematics	39	58	3	27	71	3

Table 6.2 Summary of APP progress analysis between assessment rounds – 2007/8

	Year 3						Year 4					
	Jan–Mar			Mar–May			Jan–Mar			Mar–May		
Per cent	same	up	down	same	up	down	same	up	down	same	up	down
Reading	42	52	5	41	57	2	32	55	13	40	56	4
Writing	40	57	3	51	44	5	32	62	6	36	59	5
Mathematics	35	62	3	35	60	5	39	59	2	32	63	5
	Year 5						Year 6					
	Jan–Mar			Mar–May			Jan–Mar			Mar–May		
Per cent	same	up	down	same	up	down	same	up	down	same	up	down
Reading	36	58	6	31	64	5	30	67	3	38	50	11
Writing	44	52	4	44	52	4	30	66	4	38	55	6
Mathematics	40	56	4	37	57	7	27	71	2	43	53	4

More importantly, in addition to the simple level outcome, schools can draw on the detailed qualitative outcomes from APP to track not just an improvement in level but to track the developing skills and understanding revealed by the profile across assessment focuses.

Where the level achieved by individual pupils improved between one assessment round and the next, in the vast majority of cases this was to the next sub-level, which would be expected. However, in the case of a small number of pupils, teacher judgement changed substantially between one round and the next (the maximum was a change of four sub-levels). Taking account of the fact that teachers have been developing skill in the use of APP over time, it is likely that these

are anomalous judgements rather than true reflections of progress. There were far fewer of these 'anomalous' changes between rounds in the second year of the pilot.

Changing patterns in assessment outcomes

If APP is having an impact on teaching and learning, then one way that this might be seen would be in changing patterns in assessment outcomes. Looking at the assessment focus level distribution data in appendices 3 and 4 indicates that this might have been the case. The charts show that, as would be expected, there is an improvement in the level outcomes for all attainment targets between the two assessment rounds. However, there are indications that assessment focuses where performance was weaker in the first round tended to improve proportionately more than other assessment focuses. For example, AF1 in writing and 'Operations and relationships' in Ma2 between autumn 2007 and summer 2008 both show greater improvement than the assessment focuses where performance was relatively stronger in autumn.

Teacher perceptions of progress

In summer 2008, 97 per cent of the teachers responding to the questionnaire said that APP had improved their ability to identify gaps in pupils' learning. We know from section 5 that they took action to try to modify their teaching to address any issues stemming from lack of coverage which were revealed by using APP assessment. Feedback from all sources has indicated that the patterns of strengths and weaknesses for their focus group pupils prompted them to take action on weaker areas.

'Any weak areas were focused upon and next steps identified from the assessment focuses.'

'Using more focused comprehension work to try to raise inference skills'

'Planning included more work on connectives and linking paragraphs and vocabulary as these were weaknesses identified in last round.'

Chart 6.1 from summer 2007 shows that teachers working in mathematics, for example, were conscious of attempting to use what they learned from APP to improve differentiation for their pupils and that, for both English and mathematics, outcomes were being used to identify next steps.

Chart 6.1 Use of assessment outcomes – summer 2007

Q15 How have you used MCP assessment outcomes?
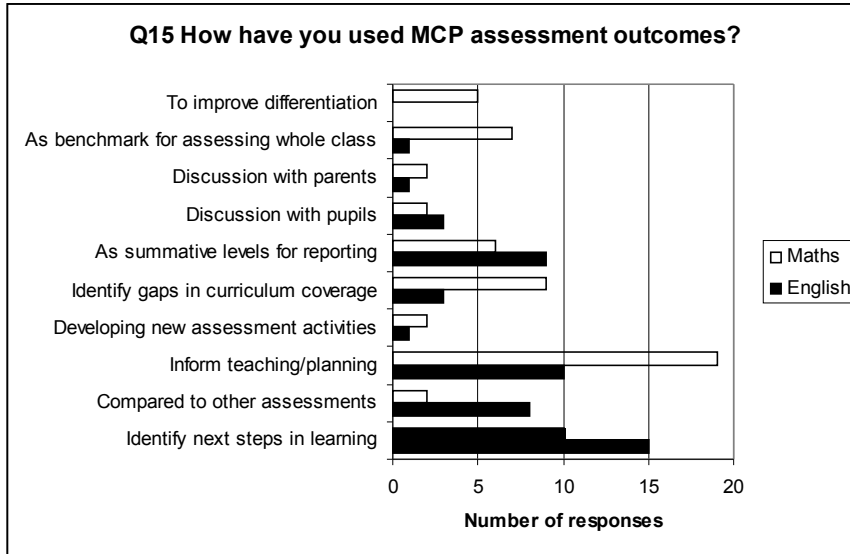
Ultimately, improvements to teacher assessment might be expected to result in improvements in the progress made by pupils. Teachers and headteachers have been asked about their perceptions of the impact of using APP on pupils' attainment and progress and have also volunteered comments about this.

> 'The pupils have benefited enormously from the project and it has increased their confidence in mathematics.'

> 'I've been amazed by how much they can do if given opportunities and challenged.'

Chart 6.2 shows that around 60 per cent of teachers felt that they could already see an impact on pupil attainment after using APP for three terms.

Chart 6.2 Teachers' views on APP impact on pupils – summer 2007

Q14b Have you identified specific changes in what your pupils can do?
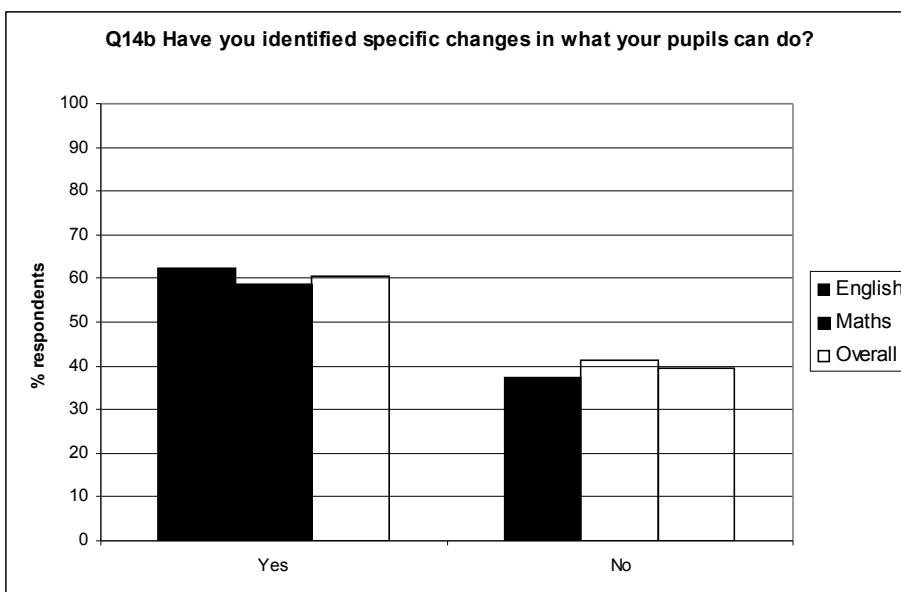
Giving details of these changes, teachers noted:

- improved inference skills

- better able to structure their writing

- more confident problem solvers

- more confident in taking risks

- better at working as part of a team

- better at sharing ideas

- take more opportunities to make decisions or access open-ended tasks

- involved in a wider range of reading activities

- better aware of what they can do – linked to confidence levels and self-esteem

- better able to explain – in all aspects of reading

- better able to explain their calculation strategies

- produce more personalised work.

At the same time point, headteachers were asked if they were aware of any evidence that pupils were making better progress:

Chart 6.3 Headteacher view on pupil progress and APP – summer 2007
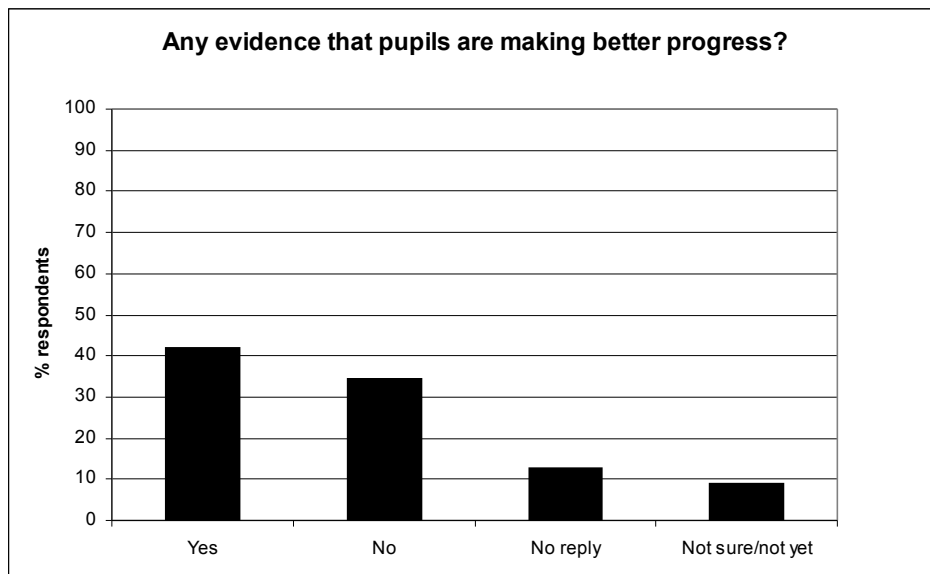
At this stage, just under half the respondents were confident that APP was related to more rapid progress but only a few could refer to any specific evidence to support their view. There were four references to using tracking systems to compare actual achievement in an 'APP' class to expected achievement based on previous data; one reference to improved achievement at key stage 2 for year 6 pupils and other anecdotal comments on comparing work produced by pupils where APP was used to the work of other pupils in the same year group.

'Too early to say; initial thoughts are "Yes".'

'Impacted on reading assessment – standards have improved as a direct result.'

'More opportunities for attaining higher levels.'

Some headteachers commented on the greater self-esteem and focus among pupils or reported that their teachers were confident that the pupils were making better progress. Generally, there appeared to be an impression that APP was supporting progression but that more time would be required for the impact to become apparent.

7. Accuracy in assessment judgements

Experience in the first year of the pilot

By the end of the first year of the project, 65 per cent of teachers responding to questionnaires were confident in the accuracy of the judgements they were making using APP and 34 per cent were 'very confident'.

Chart 7.1 Teachers' confidence in APP judgements – summer 2007

Q11 How confident were you that your level judgements were accurate?
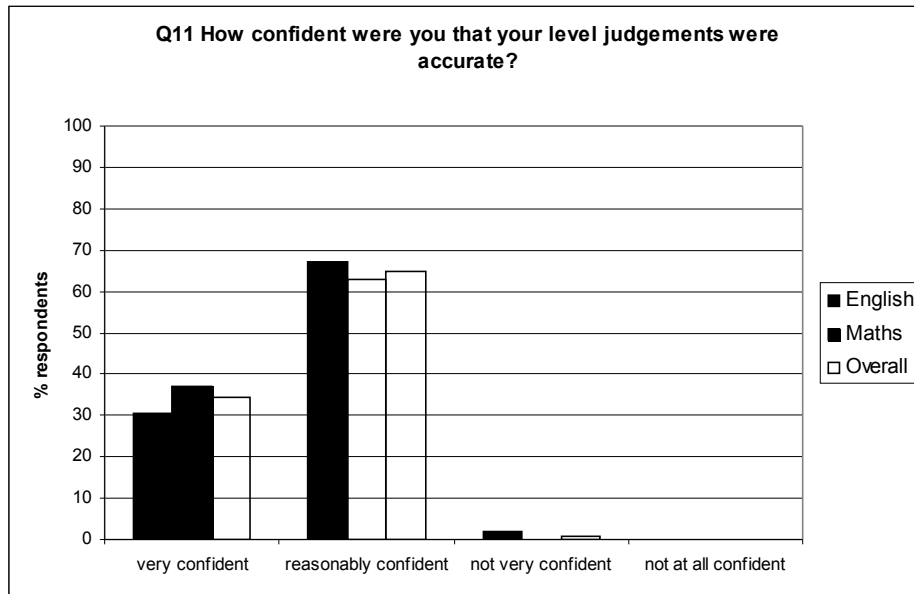
An external view of whether this confidence was justified came from the work done by the team of external moderators, through analysis of copies of the individual reports returned to teachers after each visit and from their termly summary reports to the team.

In the first year of the pilot the judgements in every school were subject to moderation and over 10,000 individual judgements at assessment focus level were reviewed. There was a different focus for moderation in the three assessment rounds so that, for example, in the first term for mathematics only Ma1 and Ma2 were moderated (although teachers themselves were asked to make and submit judgements for all attainment targets in each round). Appendix 7 includes tables which compare the moderator judgements to those made by teachers in the first year of the pilot.

Overall there was an increasing degree of agreement between teachers and moderators over the course of the year and, by summer 2007, moderators were confirming judgements for almost all the assessment focuses in more than 90 per cent of cases. Where there were differences between teachers and moderators, mathematics teachers were more likely to underestimate the level. There were concerns from moderators that there was often a reluctance to consider whether evidence pointed towards the level above and sometimes a reluctance or failure to credit evidence of attainment. In English, differences between moderators and teachers were more likely to arise because a teacher had overestimated the level for an assessment focus. Comments from

moderators suggested that inaccuracies stemmed from a lack of understanding of some of the assessment focuses and that sometimes these were related to lack of sufficient evidence to show attainment across a range of activities.

In their earliest attempts at assessment, teacher judgements made across the assessment focuses to derive an overall level were far more likely to be considered inaccurate by moderators than their judgements for the individual assessment focuses. Many teachers were not following the process set out in the guidance materials (the 'flowchart') on how to make this overall judgement but instead were just using the visual impression from their completed assessment grids to 'come up with' the overall level. This reflected a view, expressed by teachers at some of their meetings, that the final level outcome was of less importance to them than the process of evidence review and identifying strengths and weaknesses in the individual assessment focuses. There are implications in this view for how APP is used in the longer term. In the pilot work, the importance of using of the guidance and the standards files, which also help to support teachers in making accurate judgements, was stressed both at local authority meetings and by the moderators in their subsequent visits, ensuring that the whole process was followed correctly. In the second and third rounds there was less reluctance to commit to these overall judgements, and in summer 2007 they were confirmed by moderators as follows:

- Ma1: 90 per cent

- Ma2: 98 per cent

- Ma3: 93 per cent

- Ma4: 92 per cent

- reading: 79 per cent

- writing: 69 per cent.

In the plans and guidance provided in the autumn term 2006, schools were asked to organise in-school standardisation and in-school moderation activities for each assessment round to help promote accuracy, and advised that these activities should take place before the visit of the moderator. It was not always clear how consistent schools were in following their plans. Teacher questionnaire returns suggested that in-school standardisation was taking place in around two-thirds of schools and in-school moderation in 80 per cent or more. However, moderators did not always detect evidence that this was the case and it was also obvious that schools varied in their understanding of the purpose of standardisation and moderation. In the earliest days of the project, in-school moderation was quite often simply a joint discussion of evidence and a shared judgement. Where moderators were aware that teachers, with the support of senior managers, had worked hard

to develop their understanding and use of standardisation and moderation, they reported that there was a consequent improvement in the accuracy of judgements over the first year.

Experience from the second year

In the second year of the pilot, the experience of moderation was very different for most schools. Schools took part in moderation activities which were organised within their local authorities and, in some cases, these activities included a significant number of schools which were not part of the main key stage 2 pilot group. In terms of the moderation process, a key difference was that teachers were involved in the moderation of the work of others, making moderation an active process based on mutual challenge, rather than the more passive process of a one-to-one visit. Another difference was that, in some of the models, the teachers 'handed over' the work and judgements they were presenting for the scrutiny of others and were not present at discussions to justify or add to the information available to those moderating the assessments. Finally, unlike the first year, there was no guarantee of moderation in each of the three terms; the number of experiences of moderation depended on the arrangements made by the local authorities.

Other features were similar to the arrangements in the first year. The focus for moderation varied from term to term and was determined by the local authority in consultation with the moderation team supporting the activities. There was, for example, very little moderation of reading in the first term but almost no moderation of writing in the summer term as the focus moved over to reading. Every school received a written feedback report and copies of all collections of work submitted for moderation were retained so that a sample could be used for a 'national' accuracy check.

Many of those whose judgements were reviewed in the first round of moderation were therefore completely new to the use of APP and, because of the changing focus, their second experience of moderation was often their first for that particular attainment target. With this in mind, the accuracy of judgements in the second year has not been directly compared to those from the first year. Moderation in the second year has been considered as a separate source of evidence for accuracy, gathered from processes which are themselves under development. The moderation pilot has provided insights into how support for APP assessment might be organised in the longer term through local authorities or school cluster groups.

Chart 7.2 shows the way in which the accuracy rating given in moderator feedback reports varied with the number of rounds of moderation and assessment undertaken. Interestingly, there is little difference between accuracy levels in the first and second rounds, perhaps because, as has been mentioned, even if it was a second experience of moderation, it would always have focused on a different attainment target. However, those schools which had three rounds of moderation were inevitably re-visiting subject areas with which they had developed familiarity in terms of assessment and there was evidence of a significant shift towards greater accuracy, as measured by the confirmation of judgements through moderation.

Chart 7.2 Moderator view of accuracy – second year of the APP pilot

Moderator feedback on accuracy of teacher judgements in second year of the pilot
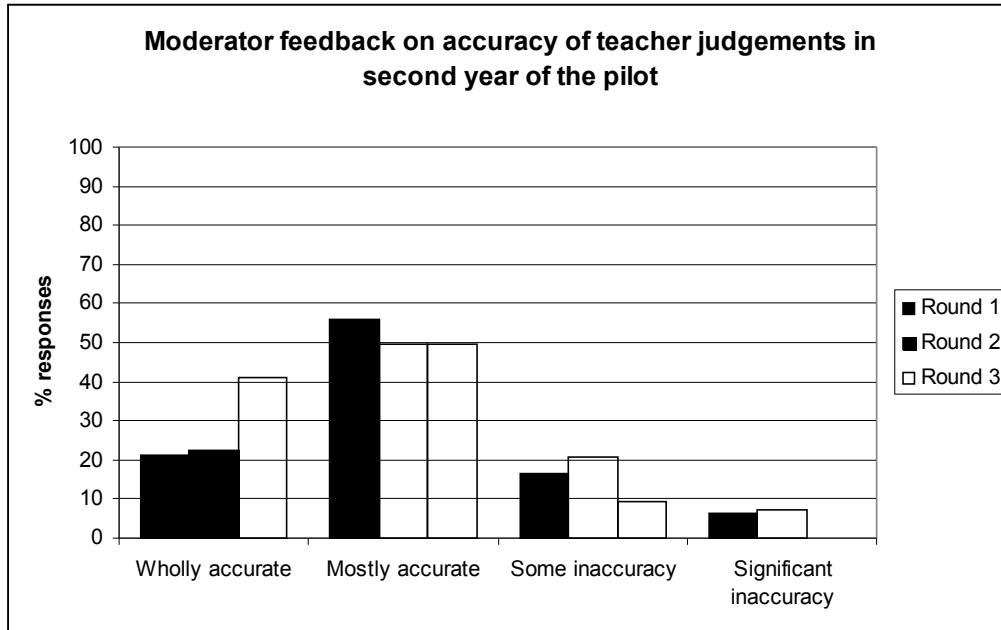
This increase in accuracy with growing familiarity reflected the experience in the first year. By the final term, judgements made by this group of teachers, from both pilot and non-pilot schools, were being confirmed as correct in 75 per cent of schools for mathematics and 76 per cent of schools for English. Moderators have commented on the substantial improvement over the year in terms of the sound evidence base for judgements with increasing awareness of the process among the teachers with whom they worked during this year.

Teachers have generally been very enthusiastic about their involvement in a range of group moderation activities, despite their initial apprehension about having to comment on the judgements of colleagues and about having to face possible challenges to their own judgements. In their evaluation feedback forms, the overwhelming majority of teachers considered the activity to be effective or very effective in improving the quality and consistency of their judgements, regardless of the particular model experienced. They valued the opportunity to check the consistency of their judgements across schools and appreciated the chance to share ideas about best practice in assessment. They were reassured when their judgements were confirmed and took away important insights into standards where they were not confirmed. Objectivity was considered to be maintained because of the structured nature of the APP criteria and the ability to refer to the standards files, as well as being able to refer to QCA or local authority moderators for advice.

'Time spent on moderation is invaluable.' (headteacher)

8. Manageability of APP assessments

Before considering aspects of manageability, it is worth noting that in piloting APP, there were inevitably features which were not an integral part of the APP process as it would be implemented in non-pilot schools, but which existed to support the experimental and developmental nature of the work. The problem with this, as in any pilot or trial, is that the participants see their whole experience as representing APP. So, for example, the termly data collections at fixed times, which probably drove the timing of periodic reviews, rather than the need to check progress, were seen as part of APP. Moderation on a termly basis, however helpful and supportive the process was, would never be so frequent in a wider implementation, but for many of the pilot teachers every round of periodic assessment was subject to a follow-up scrutiny. The process of moderating APP assessments was also in the course of being developed within the pilot. The materials themselves were in a state of development for the first year. Some concerns from teachers about manageability may have related to misunderstandings about what was 'for the pilot' and what was essential for APP.

> 'We found this process too time consuming and, being a small school, we have found it extremely hard to get the data in by the early deadlines, as it doesn't fit with the time that we collate our assessment.'

However, all teacher concerns about manageability highlight important issues which need to be taken into account as APP is further developed and extended to other schools.

Time

Throughout the two years of the project, the biggest problem for teachers has been finding the time for the processes involved in APP. The first term was extremely challenging. Teachers had to attend training, plan for and carry out standardisation and moderation with colleagues in school, complete pupil assessments (looking back to find evidence) using criteria with which they were unfamiliar, return data collection forms and, finally, prepare for moderator visits. This all took place in the space of a very few weeks at the busiest time of the year for primary schools. Their professional approach and commitment to this demanding schedule was impressive and appeared to reflect their enthusiasm for being involved in a new approach to assessment.

All three questionnaires to the pilot teachers in the first year of the project asked how long it had taken to complete the assessment guidelines for each pupil in the previous assessment round. Average times quoted by teachers for this aspect of the process are set out in Table 8.1.

Table 8.1 Time taken to complete APP assessment guidelines – 2006-2007

Time point	Time taken to complete APP assessment guidelines – minutes per pupil		
	English APP	Maths APP	Overall
First assessment round	74	53	64
Second assessment round	51	37	44
Third assessment round	43	36	40

The time taken for this part of the process decreased with time. Teachers themselves had predicted that it would, as they became familiar with the process and had taken care to ensure that they did not have to 'trawl back' for evidence. It is worth noting that these average figures reflect a wide range of values quoted by individual teachers. In the second assessment round, for example, the range was from 20 minutes per pupil to four hours per pupil. Obviously, it is possible that some of the teachers had misinterpreted the question and given the time for completing the guidelines for all their focus group pupils, or that they included time spent on other related activities, but there was plenty of verbal feedback from teachers at their early meetings about the effort involved and accounts of spending weekends completing the grids.

As well as taking less time, teachers reported that using the APP guidelines became easier over time and, when asked in the second assessment round why they felt this was, most teachers referred to their growing familiarity with the assessment focuses and the APP criteria. Pre-planning for assessment opportunities and that fact that the second round was not 'starting with a blank sheet' but was effectively a check on progress made since the last round, also helped to make the process less daunting.

Chart 8.1 Teacher views on how APP assessment becomes easier over time – spring 2007
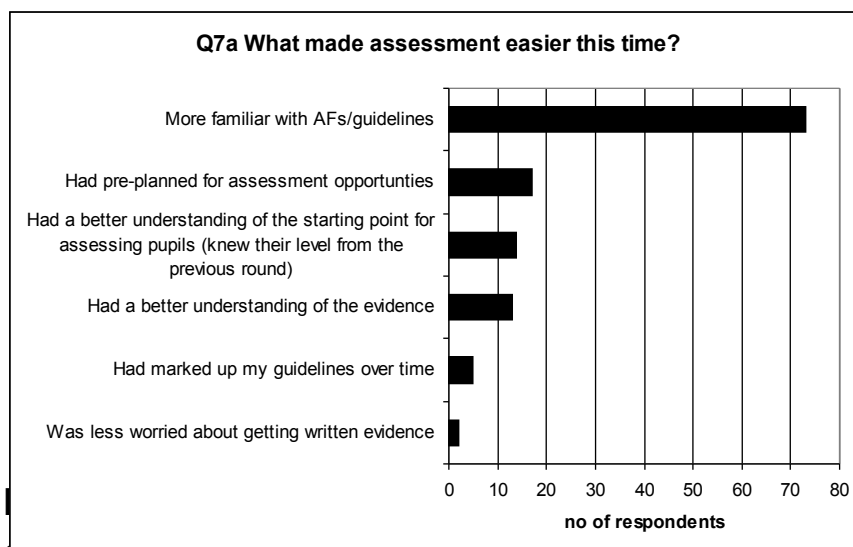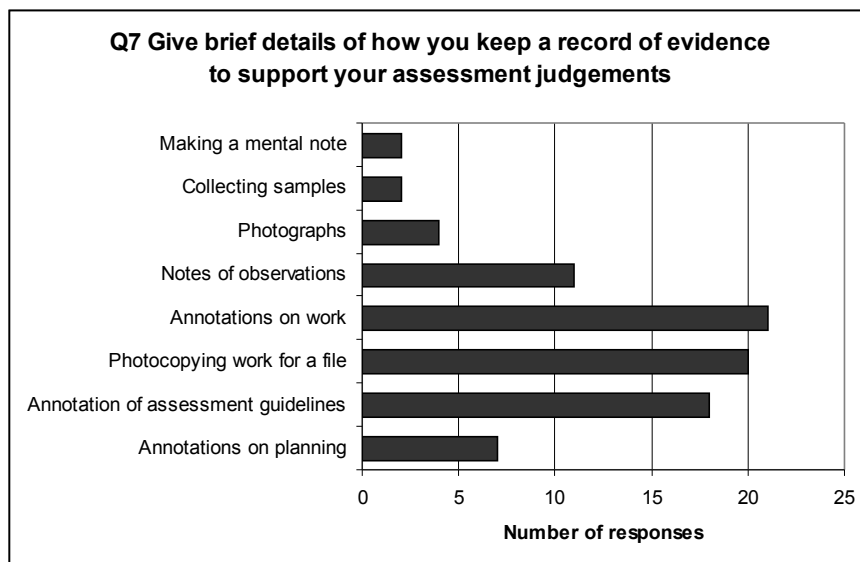

Completing the guidelines for the periodic review is the most visible part of the process but there are other aspects of APP which have to be managed and require time and effort. As discussed earlier,

teachers became aware of the need to 'capture' evidence and for many this often required them to plan specifically for such activities as observation, questioning and pupil interviews. They also needed a manageable way to record the evidence they were capturing and there was concern about whether some teachers may have been less than clear about the difference between making a brief note of evidence on the guidelines and constantly updating judgements. The first option was a sensible idea used by many teachers but the second, using a kind of 'continual assessment' tick-list, not only creates work but actually dilutes the power of APP which comes from standing back at some distance from teaching input to look at progress.

In summer 2007, the teacher questionnaire asked for brief details of how teachers were recording evidence to support their assessment judgements. Their responses were coded and grouped and are shown in Chart 8.2.

Chart 8.2 Ways of recording evidence – summer 2007

Q7 Give brief details of how you keep a record of evidence to support your assessment judgements

Category	Number of responses
Making a mental note	2
Collecting samples	2
Photographs	4
Notes of observations	11
Annotations on work	21
Photocopying work for a file	20
Annotation of assessment guidelines	18
Annotations on planning	7

Teachers were given suggestions of manageable ways to note evidence to contribute to APP assessments at the initial training and in follow-up meetings; these included the use of 'post-it' notes, notes made on plans, photographs and comments when marking workbooks. The emphasis was clearly on minimal effort for the maximum benefit to teaching and learning. Chart 8.2 shows that many of the suggestions were taken on board and there was also feedback about the use of teaching assistants to help with capturing evidence and how beneficial it was to include them in the APP assessment process.

Looking at this chart however, there was an obvious disjunct between the message given to teachers at the start of the pilot that there was no need for portfolios of work and the fact that by the end of the first year almost all teachers were regularly photocopying work and presenting 'portfolios' of evidence, photocopied or otherwise, for moderation.

In the spring term, teachers attending review meetings started to consistently comment on two aspects of their moderation experiences:

- the variation in expectations of different moderators – some were markedly more demanding than others in terms of what teachers should prepare;

- the different model of moderation experienced by teachers in English compared to mathematics.

In mathematics, the approach was that moderators expected to work through sample judgements alongside the teacher involved. They looked for evidence in written work but also probed teachers for anecdotal and more ephemeral evidence through discussion. This had several important implications. It reinforced the value placed on what teachers know about pupils from everyday observation as well as on recorded evidence and it allowed moderators to make full use of this kind of evidence in arriving at a judgement. In English, there was more variety in the approach of individual moderators but it was often the case that moderators were given collections of work (the 'portfolios') which they reviewed alone first and then met with the teachers for discussion. Where sufficient time with the teachers was available, the full discussion supported the same outcomes as the model in mathematics but in many cases the discussion time was not extensive and in their commitment to ensuring that judgements could be effectively moderated, moderators were pushing teachers to include more contextual information and more annotation in the collections for moderation.

At the end of the first year, teachers were asked directly how long it was taking them to prepare for their external moderation visits. The average time quoted by teachers using English APP was 79 minutes and for teachers working on mathematics it was 61 minutes. The range quoted was from 10 minutes to three hours. Moderation was usually based on the judgements for two sample pupils. From observation of moderation visits and through moderator reports, it was clear that in some cases teachers were producing extensive, highly annotated collections with lots of contextual information, almost rivalling the detail set out in the standards files. The effort involved must have had a significant impact on their view of manageability.

In the second year, teachers in general attended fewer moderation events. In some local authorities, there were two rounds of moderation; in others there were three. Not all schools sent teachers along to every moderation event and different teachers from the same school often took turns in attending. In terms of the process, teachers were quite clear in their feedback that they found the experience of taking part in these moderation meetings with colleagues from other schools extremely beneficial but once again the time required to prepare caused them great difficulty. Particularly in the group models where teachers do not work directly on the collections they present, there is pressure to include contextual information and annotations on work so that the moderation outcomes will be as fair and accurate as possible. This is in addition to completing the two covering forms required for each collection. Some of the models required a 'spare collection' to be prepared, even though these were rarely used in the meetings and to teachers this was just another chunk of photocopying for no obvious benefit.

In the very public context of such meetings, teachers do not want to produce a collection which is seen as less adequate than others. This tension between, on the one hand, what is required to support teacher assessment in school, giving accurate information about strengths and weaknesses, which can be added to by teacher knowledge of what that pupil does and says every day in class, and, on the other hand, what is required for 'outsiders' to be reassured that assessment judgements and processes are robust and in line with national standards, is still unresolved for many teachers. At the end of the project, when asked to identify drawbacks with APP, there were several comments about the preparation for moderation including:

'Too much paper-based evidence is required.'

'Time taken to collect and annotate work ready for moderation.'

'I feel a little disheartened over evidence – teacher professionalism meant to be trusted but still there are hints at reams of evidence.'

Time to take part in APP related activities

The requirements for time outside the classroom for APP related activities, including training new teachers, in-school standardisation, in-school moderation, collaboration with colleagues and attending external moderation, have proved burdensome for many teachers. There has been great variation in the extent to which schools have supported their teachers by allowing and funding time for these activities. This was particularly evident in autumn 2007, when many schools were taking on the challenge of introducing new teachers to APP and, in many cases, extending the use of APP into new classes, a new subject or across the whole school. In their questionnaire responses in December 2008, teachers overwhelmingly identified having insufficient time as the most challenging aspect of APP at that stage in the pilot.

Chart 8.3 Challenges in APP – autumn term 2007
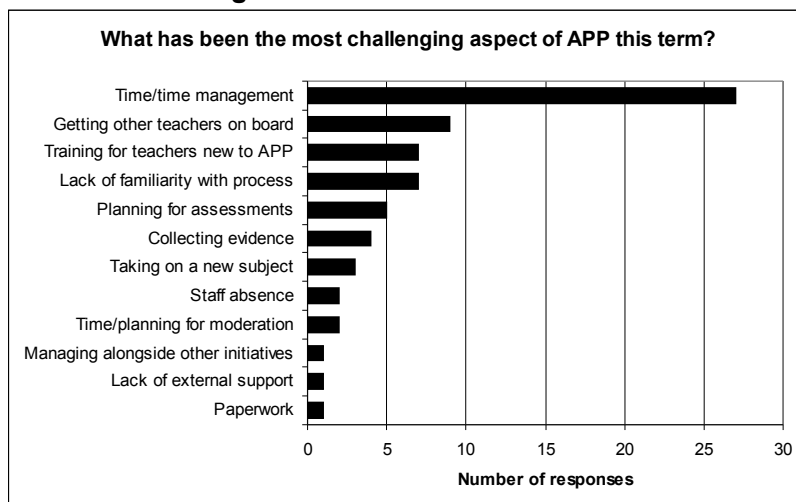
'This is a great idea and system, but it has been too much too soon for an overworked staff in a one-form entry primary. Not enough release time.'

Using APP for English and mathematics

In the initial set-up for the project, there was an expectation that while a school might use APP for English and mathematics in the pilot, this would not necessarily mean that any individual teacher would take on APP for both subjects immediately. There was some concern among teachers that this would be too much to manage while unfamiliar with the system. However, by the start of the second year, when many schools were already extending their use of APP, 50 per cent of school replies to the questionnaire in autumn 2007 showed that at least some of their teachers were then using APP for both subjects. There is no direct evidence about any particular difficulties with managing APP for both subjects, although plenty of feedback has suggested that teachers think it is best to start using APP for a single subject and build on from that. Interestingly, in the APP key stage 1 pilot many teachers took on both subjects immediately and do not seem to have had any more problems than the key stage 2 teachers using only one subject in relation to manageability.

Teachers working without support

A small number of teachers have continued to use APP in an almost isolated way since the start of the project. For them the combination of little support from senior managers and lack of local authority-coordinated opportunities to work with other schools have left them feeling very frustrated. They can recognise the potential of the system but the problems of finding time, and particularly of finding time when there has been no attempt to reduce any of the demands of existing assessment regimes, have just been too much. One teacher said how powerful she thought APP was but also:

> 'It is time consuming at the moment at school due to running a number of assessment priorities.'

One junior school at the end of the pilot simply identified one simple drawback with APP:

> 'Not enough time to implement it.'

However, the situation in many other schools was very different and where the whole school had become involved it was obvious that organisational changes were supporting APP, particularly through using regular staff meetings for in-school standardisation and moderation activities and where decisions had been made to replace some other forms of assessment with APP. In their 'top tips' for teachers new to APP, included in full as appendix 9, teachers at the end of the project gave consistent messages that managing APP becomes far less problematic with time.

> 'Bear in mind that the process does become quicker and more efficient with time and experience.'

> 'Don't be put off by the paper. It gets easier and makes sense so don't give in too quickly.'

> 'Don't be overwhelmed by the amount of work – over time it becomes much more manageable.'

They were also convinced that the effort was justified:

> 'Don't give up – once you are used to the process the benefits for pedagogy and practice are clear.'

9. Integrating APP into school assessment practice

From the outset, there were some schools who were looking to APP as a possible replacement for their current assessment regimes.

> 'If we had more proficiency in assessment it would mean that we could test the children less.'

Others were interested in enhancing teacher assessment, seeing APP as another tool which might be suitable for use alongside other systems. There were also those just interested in being part of a national pilot to keep up with new developments.

APP is of course an assessment system which schools can opt into and there was never any instruction or requirement or even guidance as to how schools should go about moving from the pilot activity to making APP integral to their assessment policies. This section looks at some indicators from teacher and headteacher feedback about the ways in which APP has been embedded at class and school level and the barriers which have been encountered.

Assessment of a focus group of pupils

The pilot requirements were participation by two teachers, each carrying out detailed assessments for a focus group of around six pupils (see annex on page 154). The rationale presented to teachers was that this detailed assessment would help them to develop the knowledge and skills about the APP criteria which they could then 'apply' to the whole class. There was discussion at training events that the assessment of the focus group pupils might act as a kind of sampling mechanism, with the selected pupils representing the range of ability across the class. Unfortunately in the first year, some teachers tended to choose a group of pupils with ability clustered around the level 2/3 or level 3/4 borderline and it was therefore hard for many of them to visualise how this sampling might work. Even in the second year, when many teachers did select a representative focus group, there was often a feeling of unease that the attention given to the focus group of pupils, while giving valuable insights into teaching and learning in the class as a whole, was of particular benefit to only a small number of pupils. When teachers became aware of the potential of APP assessment to diagnose individual strengths and weaknesses and to personalise next steps for learning, they wanted all their pupils to benefit from it:

> 'We feel strongly that each child has a right to an ongoing assessment every year.'

In most schools teachers have to provide regular assessment levels to feed into school monitoring systems and they need a mechanism for producing these for all pupils. There has been no central training or modelling of how to use detailed assessment outcomes for some pupils to produce reliable assessment for all pupils in the class. In the absence of guidance, some teachers have obviously developed their own approaches; others have not and remain uncertain that this could be effective.

We already know that managing the detailed assessments for the focus group pupils can be demanding in terms of the time it takes. There are teachers who cannot see how their detailed assessments of a few pupils can inform assessment for all, and who consider it to be impossible to manage detailed APP assessment for all pupils:

> 'You can only use APP in small groups. Assessing the whole class would not be manageable.'

> 'How can you identify individual gaps if the assessment guidelines are not for each pupil? If the assessment guidelines are for each pupil then I see this as unmanageable.'

> 'If we rolled it out to more children can you imagine the size of the portfolio.'

Other teachers appeared to be quite comfortable with the idea of the 'focus group' approach and had very sensible and pragmatic suggestions for the best way to select them:

> 'Choose the child you know least well from each of your table groups.'

> 'Select pupils with good attendance.'

> 'Change the focus child from your table group every half term.'

A third group of teachers were convinced of the value of APP, wanted it to be used with all pupils and see this as being manageable, once familiarity and expertise has been developed.

> 'I track APP reading for every child in the class as I don't feel that doing a sample group is sufficient to give individual actual levels.'

> 'We will use it as the school's assessment tool in every year with all children.'

At the end of the first year of the pilot, most headteachers responding to the questionnaire (93 per cent) disagreed with the statement that the assessment of a focus group was insufficient to give a view of progress in the whole class. This was in contrast to their teachers' perceptions that using assessments with a focus group of pupils to inform other assessments just 'would not be allowed' outside the pilot situation. There is still confusion around the 'best' way to use APP for whole class assessment; some schools are perfectly able to develop their own solutions, whereas others require guidance.

Replacing other assessments

At end of the first year of the pilot, 76 per cent of the headteachers completing questionnaires indicated that APP could replace some of their existing assessments. Their explanations for this were coded and grouped and are set out in Chart 9.1.

Chart 9.1 Headteacher comment on replacing other assessments with APP – summer 2007
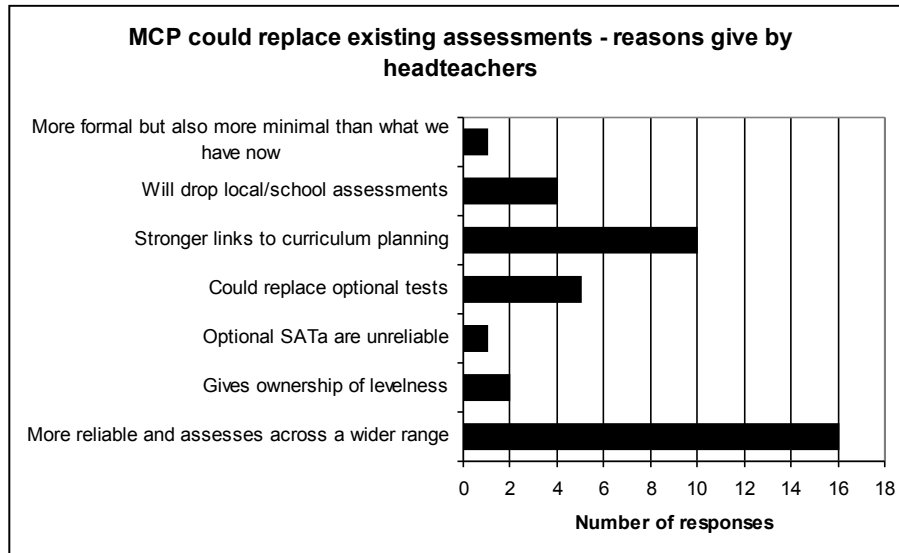
'Yes, they are thorough, moderated, credible and shared with colleagues.'

'Will replace endless end-of-term assessments.'

At the end of the second year, responses from teachers indicated that, in 38 per cent of schools, headteachers had translated this intention into action. The assessments replaced were mostly the optional national curriculum tests:

'Teachers really notice how inadequate these are for assessment now that we are more confident with APP.'

Others were rather more reluctant to drop these established tools:

'May have less need for optional tests but while end of key stage 2 tests remain the optionals are needed as preparation.'

There were some references to the fact that although optional tests were still in use, there had been a change to their status, so that they provided confirmation of teacher assessment, being treated as just another piece of evidence. The relationship between one-off tests and APP periodic assessment is explored further in section 11.

Timing of APP assessments

In the second year of the pilot, there was more opportunity for schools to 'take control' of the organisation of APP in their schools, and in the final questionnaire teachers were asked how frequently they were using APP for periodic review of progress. Seventy per cent of respondents said the review was taking place on a termly basis and 25 per cent were carrying out assessments

half-termly. There were some comments that it was being used more frequently by some individuals and this is an indication that the principle of APP as a tool for periodic review and not as a continuously updated tick-list is still not universally clear.

It seems likely that timing at the end of the project was related to school requirements for progress monitoring and 83 per cent of the questionnaire respondents said that APP was informing the teacher assessments used to feed into their school tracking systems.

Using APP with existing systems for tracking progress

In summer 2007, headteachers were asked to describe their existing assessment systems and their responses were then matched to their views on whether or not APP would fit well with these. The resulting table is included as appendix 10. Interestingly, headteachers with ostensibly identical existing practice could hold completely opposite views on whether these were an appropriate fit with APP, suggesting that attitude and perception of the added value of APP are critical and not the practical challenges of any particular regime. From the outset, headteacher support has always been vital and the schools where APP has had most impact and been seen as successful have been those where either or both the headteacher and senior leaders were involved closely in the process and have planned for sensible, staged implementation and have not insisted that APP is 'done on top of everything else'.

10. APP as a basis for communication

With parents

Teachers' use of APP as a basis for discussion of progress and next steps with parents was relatively limited within the pilot. There were questions on this at three time points (spring 2007, summer 2007 and summer 2008) and on each occasion only a minority had made use of the APP outcomes for this purpose. The proportion did increase over time, from 23 per cent of respondents in spring 2007 to 41 per cent in summer 2008. Those who had used it found it extremely valuable, allowing them to have detailed conversations with parents so that they could understand specific strengths and weaknesses and see where they might help their children to make progress. There were references to how useful it was to have a visual representation which, through the use of coloured highlighters for example, could show progress and targets quite clearly. Some teachers felt that it allowed them to demonstrate their detailed knowledge of the children very effectively.

> 'Teachers were able to clearly identify why children were still sitting at level 2, for example punctuation, while having many level 3 features – parents can then support them at home.'

> 'This is still in its early days. We are trying to educate our parents so that they understand the steps the children need to take to progress. However, this is ongoing. We are doing this through individual targets – these have been aided by APP.'

> 'Teachers have more to say to parents.'

There were other comments that APP outcomes had helped in report-writing and that reports could be more meaningful and personalised.

Given the positive comments from those teachers who have used APP, it might seem surprising that its use has not been more widespread. However, the 'focus group' approach has been followed quite strictly in many of the pilot schools and teachers feel uncomfortable in a lot of cases with apparently 'knowing' much more about their focus group pupils than about others in their class. Having a highly detailed knowledge of a few pupils and how this can help others in the class at an individual level (rather than in terms of general messages about curriculum coverage, for instance) remained an unresolved problem for many teachers throughout the pilot.

With pupils

At the end of the first year of the pilot, just under 70 per cent of headteachers responding to the questionnaire considered that pupils were more closely involved in their own assessment as a result of the use of APP:

> 'Pupils have taken ownership of their own learning; they know where they are, where they need to go and can find out how to get there.'

Teachers were asked about this in spring 2007 and again in summer 2008. After their second round of assessments, 45 per cent of questionnaire respondents said that they had made changes to involve pupils more. Of the 50 per cent who said that they had not done so, many (29 per cent of all

respondents) went on to explain that this was because they were already implementing features of assessment for learning practice prior to being part of the pilot and mentioned techniques such as personal target setting, peer assessment and the use of traffic lights.

Chart 10.1 Involving pupils in their own assessment – spring 2007
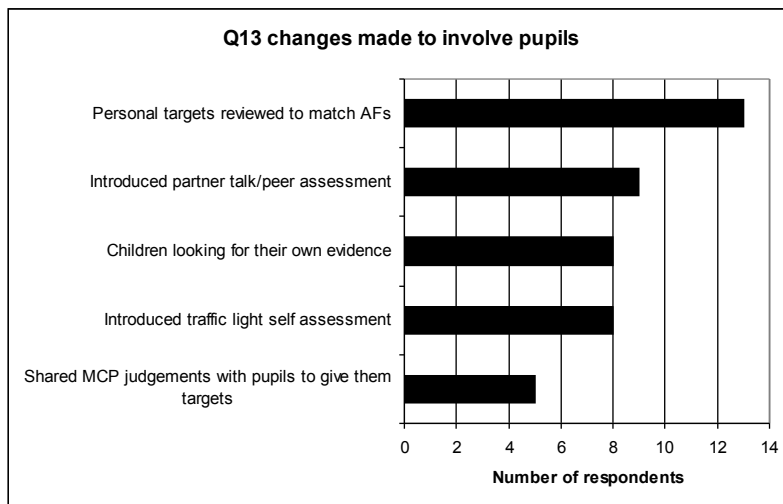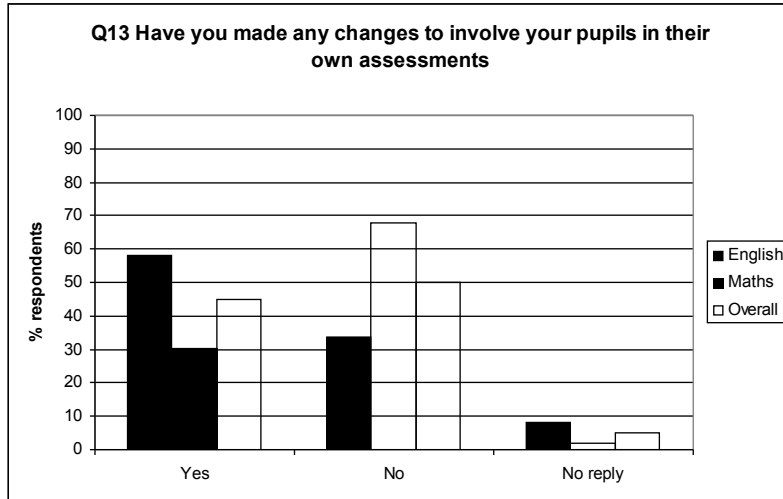
At the end of the pilot, teacher responses echoed those from the first year with only 47 per cent saying that this aspect of assessment had been improved by APP. As before, there were comments that this aspect of assessment practice was already established. Confirmation of this came from the collections presented for moderation which often included examples of peer- or self-assessment. During teacher meetings there was often reference to the self-assessment technique of 'two stars and a wish' and how useful this had been for many teachers as a source of direct evidence for their APP assessments.

With other teachers

Through attendance at moderation meetings, it has been possible to observe many instances of the way in which APP facilitates professional dialogue between teachers. The assessment criteria

provide a common language in which to describe and explain the details of performance in reading, writing and mathematics. The structured nature of the criteria and their link to nationally agreed standards in the form of the standards files allow teachers to challenge their own and others' assumptions about the implications of what their pupils can do, as well as providing them with the reassurance that they are making sound assessment judgements themselves. There have been many references to how useful it is or would be to receive and pass on completed assessment guidelines as pupils move between classes.

The change between the rather hesitant grasp of assessment focuses and NC level descriptors demonstrated at the very first training meetings and the confident, competent and informed discussions in the final round of moderation meetings in summer 2008 was striking. Through the work of some local authorities in the second year, teachers with experience of APP in the first year were developing their skills further so that they could take on the role of group leader for moderation activities. Where this was observed it was extremely effective, giving ownership to teachers not only of the APP assessment process but of responsibility for ensuring that the standard of assessments is consistent among groups of schools. The objective application of the criteria to determine a level and the ability to refer to the standards files help to diffuse and manage any awkwardness when there are differing viewpoints on the accuracy of judgements.

11. APP and national curriculum tests

Most of the pilot schools have historically made extensive use of the QCA optional national curriculum tests in years 3, 4 and 5 to support their monitoring of progress within key stage 2. All the pupils in the APP data collection sample will, of course, take the statutory national curriculum tests at the end of key stage 2. We know from the evaluation of the pilot that one effect has been to reduce the extent to which the pilot schools may use the optional tests in future. This section looks at the views of teachers on their comparative strengths and weaknesses, the relationship between the outcomes of periodic assessment and 'one-off' tests, and also examines relationships between achievement at assessment focus level and achievement in a one-off test.

Teacher views

In summer 2007 teachers were asked to consider the advantages and disadvantages of periodic teacher assessment using monitoring children's progress and the QCA optional tests, which were then being used in the almost all the schools in the pilot. In general the issues identified were common to teachers working in both English and mathematics. Any differences in response between the subjects are specifically referred to.

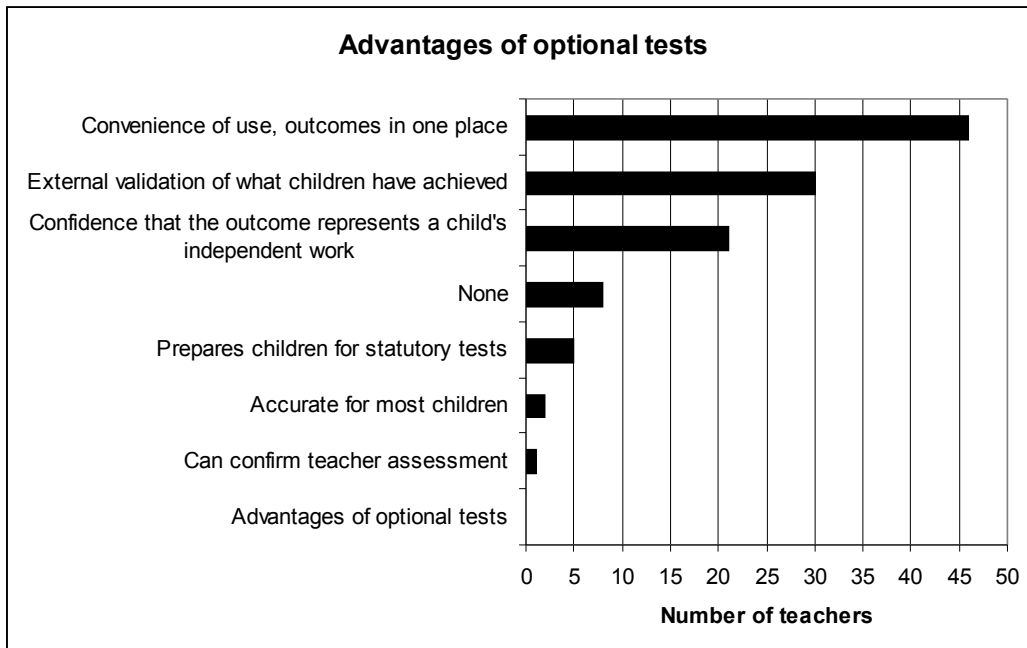**Chart 11.1 Periodic teacher assessment – advantages**
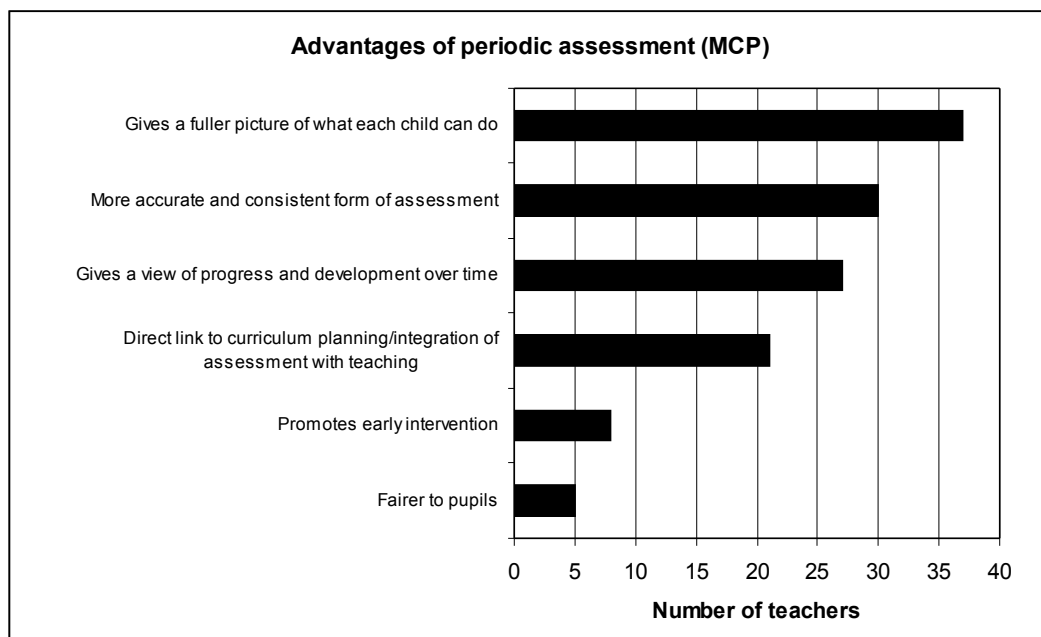
Chart 11.2 Periodic teacher assessment – disadvantages
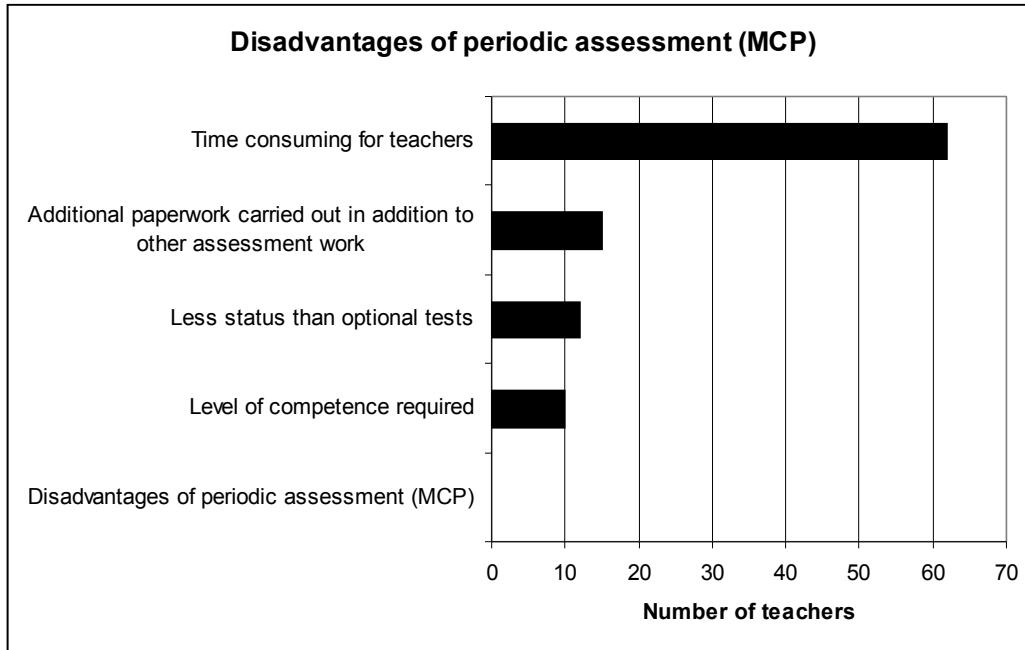
Teachers clearly felt that the key advantages of periodic teacher assessment came from the potential to form a more holistic view of what pupils are able to do which can be directly linked to planning. The drawback was the time required, particularly since it was often the case that for project schools the activity was seen as an 'extra' activity. Interestingly, all the comments concerning lower status for teacher assessment came from teachers using APP for English.

Chart 11.3 Using tests – advantages

Chart 11.4 Using tests – disadvantages

Disadvantages of optional tests

Category	Number of teachers
One off assessment carried out at a single time point	~68
Anxiety experienced by children	~17
Less accurate than teacher assessment	~13
Some children respond very badly to test environment	~12
Lack of direct relationship with planning	~6

Number of teachers
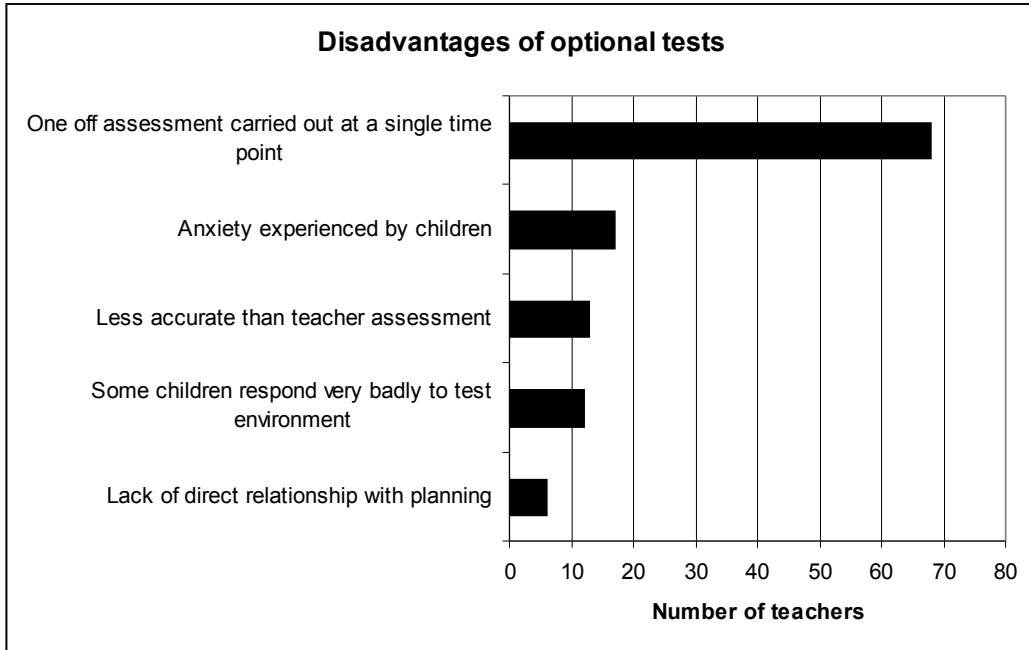
The advantages identified for test use were either related to ease and convenience of use for teachers or to a view that test use will give a recognised and accepted assessment of work which has been done independently by pupils. In terms of the quality of the assessment outcome, there were concerns that a test can only ever provide a snapshot that may not always be representative of true' achievement. There were also some concerns about the effect on children. In their questionnaire responses teachers said:

Positive features of APP:

'More meaningful and relevant.'

'Keeps you focused on progress.'

'More time to teach, less testing.'

BUT

'You can focus too much on assessment.'

Positive features of optional tests:

'Teachers are judged on them.'

'Satisfying to receive an external mark.'

'A national benchmark.'

BUT

'Taken too late for action.'

'You don't get a sense of the journey the children have come on.'

'Take time away from teaching.'

'Tests are focused on AF2 and assessment focus.' (reading)

'There's no Ma1.'

Clearly, where schools have made the decision to drop the use of the optional test, they have seen APP assessments as giving them more valuable and timely information. Other schools recognise that both types of assessment have benefits. By and large, the benefits which teachers associated with APP were related to the direct connection to teaching and learning, and the benefits of the optional tests were related to status and their link to national standards. The disadvantages of APP were all related to management and the disadvantages of the tests were linked to the limitation of what teachers find out about their pupils.

Comparing outcomes from different assessments

In the first year of the pilot there were only a very small number of pupils in year 6, so in summer 2007 it was not possible to carry out any type of comparative analysis between APP judgements and levels achieved in the end of key stage 2 tests. However, in summer 2008, the pupil sample for year 6 was included in the comparative analysis.

This comparison was carried out in two ways. Straightforward frequency analysis of level outcomes was used to create the summary in tables 11.1 to 11.6. Note that the optional tests for mathematics do not provide a measure of achievement for individual attainment targets so only the calculated mathematics overall levels can be used in the comparison. Tables of matched outcomes by year groups and in terms of sub-levels were also created; these are included in appendix 11. For the purpose of the comparison, the APP teacher judgements within a level (high, secure, low) were taken as corresponding to the calculated 'sub-level' outcomes from the tests (a, b, c).

In addition, a more complex statistical technique (multi-level modelling) was applied to test for genuinely significant differences in outcome. Multi-level modelling is a technique that allows statisticians to investigate and quantify relationships between measures.

Comparison of outcomes – summer 2007

Table 11.1 APP reading whole levels and optional test reading whole levels (per cent and count)

Reading level	Assessment type	
	APP reading	**Optional reading test**
2	8.8% (43)	7.0% (34)
3	53.3% (260)	31.1% (152)
4	37.5% (183)	54.5% (266)
5	0.4% (2)	7.4% (36)
Total	100% (488)	100% (488)

For reading, the results of the multi-level model showed no significant difference at level 2 but pupils were much more likely to attain level 3 from the APP assessment than they were to attain level 3 on the optional test. Pupils in years 3 and 4 were much more likely to achieve level 4 in the optional tests than they were to be judged as level 4 in their APP assessment. In year 5 pupils were significantly more likely to attain level 5 in the optional test.

Table 11.2 APP writing whole levels and optional test writing whole levels (per cent and count)

Writing level	Assessment type	
	APP writing	Optional writing test
2	11.9% (57)	16.1% (77)
3	60.2% (287)	49.1% (234)
4	27.5% (131)	32.3% (154)
5	0.4% (2)	2.5% (12)
Total	100% (477)	100% (477)

Overall there was no significant difference in the probability of achieving level 2 in both assessment types. Across all year groups, pupils were less likely to attain level 3 in the optional tests than they were in their APP assessment. There was no significant difference in the likelihood of pupils achieving level 4 in the two assessments but achievement at level 5 was much more likely in the optional tests than in APP assessment for pupils in year 5.

Table 11.3 APP mathematics whole levels and optional test mathematics whole levels (per cent and count)

Mathematics level	Assessment type	
	APP mathematics	Optional mathematics test
2	14.4% (63)	21.8% (95)
3	58.5% (255)	50.5% (220)
4	26.1% (114)	25.7% (112)
5	0.9% (4)	2.1% (9)
Total	100% (436)	100% (436)

In mathematics, pupils were more likely to achieve level 2 in the optional test than in their APP assessment. With the exception of pupils in year 5, they were less likely to achieve level 3 in the optional test than in their APP assessment. For level 4, the pattern varied across year groups but overall there was no significant difference in the likelihood of pupils attaining this level from either the optional test or their APP assessment. Similarly, pupils were just as likely to achieve level 5 on both the outcomes; obviously this only applies to pupils taking the year 5 optional test.

The analysis in summer 2007 showed reasonably good agreement between the outcomes of optional tests and APP assessments for pupils in the mathematics sample at level 4 and above, with evidence of difference at lower levels such that pupils tended to achieve higher levels in the optional test than in their APP assessment.

For English, there were significant disparities at all levels for reading and at levels 3 and 5 for writing. The disparity could be explained either in terms of APP assessment resulting in a tendency to underestimate achievement in comparison to the test or in terms of test results presenting a relative overestimate of achievement compared to APP outcomes. What the analysis does not show is whether either measure is intrinsically more 'accurate'.

Comparison of outcomes – summer 2008

In summer 2008 the level distribution within the whole data collection sample was quite different to that in summer 2007 as a result of the inclusion of the year 6 pupils. In fact, the year 6 pupils dominated this comparative analysis in the second year, given that much smaller numbers of schools provided optional test results for pupils in years 3, 4 and 5.

Table 11.4 APP reading whole levels and optional test reading whole levels (per cent and count)

Reading level	Assessment type	
	APP reading	Optional/NC reading test
2	6.8% (39)	6.6% (38)
3	29.0% (166)	20.1% (115)
4	47.5% (272)	43.5% (249)
5	16.8% (96)	29.8% (171)
Total	100% (573)	100% (573)

For reading, the results of the multi-level model showed no significant difference at level 2 but pupils were less likely to be judged at level 3 in a test than they were on APP assessment. In year 4, pupils were more likely to achieve level 4 in the optional test than they were from APP assessment. Pupils from years 5 and 6 were more likely to achieve level 5 in the optional or the end of key stage 2 test than they were to achieve level 5 from APP assessment.

Table 11.5 APP writing whole levels and optional test writing whole levels (per cent and count)

Writing level	Assessment type	
	APP writing	**Optional/NC writing test**
2	10.8% (62)	11.0% (63)
3	37.7% (216)	38.6% (221)
4	38.9% (223)	42.1% (241)
5	12.6% (72)	8.4% (48)
Total	100% (573)	100% (573)

Across all the year groups, there was no significant difference between test or APP outcome for pupils achieving levels 2, 3 and 4. However, pupils were less likely to achieve level 5 in the optional or end of key stage 2 test than they were to achieve level 5 through APP assessment.

In mathematics, there was no significant difference in the outcomes from tests or APP assessment at levels 2, 4 and 5. Pupils in year 3 and 4, however, were less likely to achieve level 3 in the optional tests than from APP assessment, while pupils in year 5 and 6 were more likely to achieve level 3 in the optional or end of key stage 2 tests than they were from APP assessment.

Table 11.6 APP mathematics whole levels and optional test mathematics whole levels (per cent and count)

Mathematics level	Assessment type	
	APP mathematics	**Optional/NC mathematics test**
2	7.4% (43)	13.0% (95)
3	38.3% (223)	36.2% (211)
4	40.3% (235)	36.2% (211)
5	14.1% (82)	14.6% (85)
Total	100% (436)	100% (436)

The analysis in the second year for mathematics revealed little difference between APP and test outcomes, except for the level 3 outcomes where it seems that pupils working around this level in year 5 and 6 appear to do less well in the tests than in APP assessment.

In writing, there were no significant disparities between test outcomes and APP assessments in summer 2008. However, for reading, there were differences and these suggested that pupils were more likely to achieve higher levels in the tests than from APP assessment.

Relationship between individual AF levels and optional test outcomes (summer 2007)

At the end of the first year, a 'Kappa' analysis was carried out to explore the strength of the relationship between individual assessment focus levels, as judged in APP, and the optional test outcomes for individual pupils. The aim was to explore similarities and differences between the optional test coverage of the skills highlighted by the assessment focuses. Very simplistically, it allows us to see whether the two types of assessment are measuring the same 'thing' and to see if there is any evidence for the teacher view that the tests give a narrower view of achievement in a particular subject.

The Kappa analysis classifies relationships in order of increasing strength as poor, slight, fair, moderate, substantial and almost perfect.

Table 11.7 Reading – agreement between assessment focus outcomes in round 3 and level achieved in the optional tests – summer 2007

Assessment focus	Round 3
AF1 Use a range of strategies including accurate decoding of text to read for meaning	Slight (0.175)
AF2 Understand, describe, select or retrieve information, events or ideas from texts and use quotation and reference to text	Fair (0.28)
AF3 Deduce, infer or interpret information, events or ideas from text	Fair (0.39)
AF4 Identify and comment on the structure and organisation of texts, including grammatical and presentational features at text level	Fair (0.217)
AF5 Explain and comment on writers' use of language, including grammatical and literary features at word and sentence level	Fair (0.217)
AF6 Identify and comment on writers' purposes and viewpoints and the overall effect of text on the reader	Slight (0.123)
AF7 Relate texts to their social, cultural and historical contexts and literary traditions	Slight (0.143)

In reading, AFs 2, 3, 4 and 5 were found to have stronger relationships with the level outcome on the optional test. The other assessment focuses have very little relationship to optional test level outcome. Table 11.3 shows the attribution of questions in the current optional tests against the assessment focuses. Quite clearly, the tests offer less opportunity to look at some of the higher order reading skills (AF5, 6 and 7) and this may be one of the reasons why teachers felt strongly

that their MCP assessments gave a more accurate and more rounded view of their children's reading skills.

Table 11.8 Attribution of reading questions to assessment focuses in optional tests

AF/Nos. of questions/ text type	Y3 optional test: The hunt for the secret treasure	Y4 optional test: Antarctic adventures	Y5 optional test: Changes
AF2	Information:10 Narrative: 4 Total = 14	Information: 3 Narrative: 8 Diary: 4 Total = 15	Fact file: 9 Poem:1 Information:6 Total = 16
AF3	Narrative:11 Total = 11	Information: 1 Narrative: 3 Diary: 5 Whole booklet: 3 Total = 12	Fact file:1 Poem: 8 Information:1 Whole booklet:1 Total = 11
AF4	Information: 4 Narrative:1 Total = 5	Information:1 Diary:1 Total = 2	Fact file:1 Poem:3 Total = 4
AF5	Narrative:1 Total = 1	Narrative: 1 Diary:2 Total = 3	Poem:1 Information: 1 Total = 2
AF6	–	Information: 2 Diary: 1 Total = 3	Poem: 2 Total = 2
AF7	–	–	–

Table 11.9 Writing – agreement between assessment focus outcomes in round 3 and level achieved in the optional test – summer 2007

Assessment focus	Round 3
AF1 Write imaginative, interesting and thoughtful texts	Fair (0.343)
AF2 Produce texts which are appropriate to task, reader and purpose	Fair (0.36)
AF3 Organise and present whole texts effectively, sequencing and structuring information, ideas and events	Fair (0.39)
AF4 Construct paragraphs and use cohesion within and between paragraphs	Fair (0.293)
AF5 Vary sentences for clarity, purpose and effect	Moderate (0.442)
AF6 Write with technical accuracy of syntax and punctuation in phrases, clauses and sentences	Fair (0.308)
AF7 Select appropriate and effective vocabulary	Fair (0.375)
AF8 Use correct spelling	Fair (0.318)

In writing, the relationships between assessment focuses and the optional test level were slightly stronger in general than for reading and there was less difference between individual assessment focuses. The best 'indicator' of performance in the optional test was AF5 (sentence variety).

The mathematics optional test does not provide outcomes for the separate attainment targets, so only the relationship between assessment focuses and overall test level can be examined. Within the optional tests, the question distribution across the attainment targets follows the same rubric as the weighting of the targets in the overall mathematics level (Ma2, 5: Ma3, 2: Ma4, 1) but in the

tests, Ma1 questions are not separately attributed. Ma2 would be expected to have the greatest influence on overall test outcome.

Table 11.10 Mathematics overall level – agreement between assessment focus outcomes in round 3 and the level achieved in the optional test – summer 2007

Assessment focus	Round 3
Problem solving	Moderate (0.435)
Communicating	Fair (0.4)
Reasoning	Fair (0.354)
Number and the number system	Moderate (0.436)
Fractions, decimals, percentages and ratios	Moderate (0.438)
Operations and the relationships between them	Moderate (0.413)
Mental methods	Moderate (0.433)
Solving numerical problems	Fair (0.382)
Written and calculator methods	Moderate (0.438)
Properties of shape	Moderate (0.421)
Properties of position and movement	Fair (0.396)
Measures	Moderate (0.437)
Processing	Moderate (0.41)
Representing	Fair (0.38)
Interpreting	Fair (0.401)

Generally, the stronger relationships were with most of the assessment focuses for Ma2 (except 'Solving numerical problems'). 'Problem solving' and 'Communicating' for Ma1, 'Properties of shape' and 'Measures' from Ma3 and 'Processing' in Ma4 also had reasonably strong relationships with test outcome.

In summary, APP teacher assessment and tests give different perspectives on the achievement and ability of individual pupils. Each approach has strengths and limitations, readily identified by teachers. As a result of the different nature of the assessments themselves, the outcomes for an individual child may not always generate the same level. However, teachers can use outcomes from both types of assessment to add to their knowledge of the progress being made by an individual, as long as they have a clear understanding of the contribution made by each type. Evidence from the evaluation suggests that using APP may have helped teachers to become more discriminating and 'expert' users of the national curriculum optional tests, able to see when their use is appropriate and able to challenge outcomes with their own evidence if these are not representative of what individual pupils can do.

'If there is a difference between levels from the optional test and my assessment, I will know which is correct and be able to prove it.'

12. Situation in pilot schools at the end of the project

Questionnaires were completed in the second year of the pilot by 67 schools, and 74 schools continued to submit pupil assessment data each term. (Note that of the 67 schools, 66 could be matched to known details of local authority, original subject selection, and others.) We know that other schools considered themselves to be continuing to take part in the pilot (through participation in activities such as moderation trials) even through they 'opted out' of the data collection and the formal evaluation. Looking in detail at the current situation and attitude towards APP in these 67 schools at the end of the pilot gives a useful indicator of the extent to which APP has been 'taken on board' by schools and can provide some insights into how APP might be taken up outside the pilot.

Coverage of subjects and year groups

Of the 66 schools, 32 opted to use APP for both English and mathematics at the start of the project; 17 started with English only and 17 with mathematics only. In all cases two teachers per subject were invited to the initial central training but the questionnaire respondents indicated that in the second year, APP use had been extended in the schools as follows:

Table 12.1 Extending APP across the school

Use of APP in autumn 2006	Number of schools	Number of schools extending use to all year groups
English and mathematics	32	15
English only	17	10
Mathematics only	17	7

Often, schools had APP in place in a number of year groups but had still to extend into year 6. Schools also had the option in the second year to take on a 'new' subject. Of the 17 schools who had started with English only, five went on to introduce mathematics in the second year and exactly the same proportion of schools starting with mathematics (five out of 17) added English APP. It is worth noting that for schools introduced to APP through local authority 'early adoption' of APP, most took on both subjects.

Plans beyond the pilot

The final questionnaire asked about school plans for APP beyond the pilot. Of the responses, only six indicated uncertainty about what the plans for their school were. There were two comments from schools known to have received little support from their senior managers and local authorities that they would be continuing their own limited implementation of APP and one of these schools, which had dropped mathematics at the end of the first year of the pilot, said it would be extending English into two more year groups.

The picture in schools where there has been coherent support from local authorities and senior leaders was very different. Nineteen schools said that they would be extending APP across the whole school for both English and mathematics. Four responses referred to extension of English across the whole school, three to extension of mathematics to all year groups and two to the extension of mathematics and writing across all year groups.

'All teachers to use as a major tool for teacher assessment.'

The addition of new elements was referred to by a number of schools, depending on what they already had in place. Two schools were planning to include mathematics for the first time and four were planning to add English. Some schools had obviously taken, or were taking, a staged approach to the adoption of English and referred to adding reading or writing.

There were also references to continuing with the existing arrangements in terms of subject and year group coverage but of improving practice through measures such as embedding and developing links to planning, adding reading moderation for the first time and training new teachers.

Some responses commented that next year would see the replacement of other forms of assessment, for example optional national curriculum tests, and at least two schools have decided to use APP assessment for all children in every year group:

'Guidelines for each child in mathematics, reading and writing.'

Others were intending to follow the model from the pilot:

'All staff to have six focus pupils for moderation.'

There were comments which showed that teachers had learned from the pilot that 'starting small' would work best in their school:

'A gradual rollout, 3 children per class to start with.'

Five schools signalled their intention to extend APP into key stage 1.

Planning and moderation activities in school

In the autumn 2007 questionnaire, schools were asked whether they had an agreed process for using the outcomes from APP assessment to inform planning. At that stage 40 per cent of schools said that they had. Of the remainder, several described their own approach, acknowledging that this was not necessarily promoted at 'school level'. The teachers were asked to provide details of the process and most (13) referred to using the gaps revealed by APP assessments to inform weekly planning, so that areas of weakness or insufficient evidence could be addressed. There were two references to sessions for teachers to set targets jointly for pupils using APP assessments and one reference to progress-tracking reviews.

In the final questionnaire, teachers were asked to give details of how in-school moderation was organised in their school. We know from feedback through the moderators that this was an area of concern in the first year but that moderators felt that the situation had improved towards the end of the second year of the pilot. Responses suggested that in-school moderation had taken place in 94 per cent of cases (60 schools). In their descriptions, teachers focused on a range of different features of their in-school moderation, for example the length of meetings, when meetings took place, and the number of collections and pupils involved. However, some general pictures emerged from this collection of responses.

Eight schools provided no details of their experience of moderation in schools. In 30 cases teachers described moderation activities which included at least one joint meeting for all teachers across the school who were involved with APP. Some referred to moderation taking place within a year group or in year group pairs, for example year 3 and year 4, year 5 and year 6. One school said that it had a staged process with moderation within year group pairs followed by a cross-school session. The inclusivity of the meetings varied: in some cases only the pilot teachers were included; in others there seems to have been an effort to involve as many people as possible, perhaps because taking part in moderation was seen as a very effective way to become aware of what APP is about.

> 'One staff meeting (1.5 hours) each term for each subject, reading, writing and mathematics. Includes all full-time teaching staff, head and SENCO [special educational needs coordinator] and occasionally teaching assistants.'

There were six specific references to the involvement of the headteachers or the subject leader or coordinator, but for many others participation by senior managers was implied, because moderation had been built into whole school staff meetings or in-service training (INSET) activities (13 responses). Where teachers gave details of the number of collections and pupils considered at in-school moderation, the most common model seemed to be the review of one collection per teacher or year group, followed by some models with two collections per teacher. In some cases this would indicate that a substantial number of collections were looked at in the course of a meeting. In one case, nine collections seem to have been reviewed in the course of a meeting lasting 1.25 hours. In other schools a more considered approach had developed around a kind of 'rolling programme'.

> 'One hour termly during staff meeting time, across adjacent year groups; look at 1–2 collections each time.'

Most meetings took place out of school time but, as mentioned earlier, within regular staff meetings. In other cases the meetings appear to have been set up solely for moderation.

> 'Key stage 2 teachers meet after school for two hours; three pupil collections moderated.'

There was one reference to a single scrutineer carrying out moderation:

> 'Moderated at home by literacy coordinator.'

All other responses suggested that work was undertaken by pairs or groups. The variation in the degree of support for APP activity in schools is clearly illustrated by contrasting the above approach with the following model:

> 'In-school moderation held termly following a standardisation week where staff were given two hours of time to standardise and prepare a collections of work. Everyone's collection was then moderated in a twilight staff meeting.'

Future support for schools outside the pilot

The increasing confidence in their own assessment expertise, discussed elsewhere in this report, seems to be reflected in the pilot teachers' responses to a question in summer 2008 about what additional support they required from their local authority or elsewhere. At the end of the first year of the pilot, there was quite a strong voice suggesting that teachers felt they needed more training and support to improve their practice in APP. Perhaps this need was fulfilled by taking part in moderation activities during the second year of the pilot because by July 2008, 33 responses (49 per cent) felt that they did not need any further help and a few of these said that they were now in a position to help other schools.

When teachers wanted support it was release time, preparation time and time for moderation that concerned them (11 schools). References to local authority support varied on the basis of experience within the pilot; six schools, where local authority support had been a significant feature of the last two years, wanted the ('excellent') support to continue; and two schools where local authority support had been limited wanted their local authority to set up opportunities for cluster moderation and joint working with other schools. There were four general requests for 'more training' and then some specific training needs related to plans for APP extension, for example further support on literacy for a school about to introduce APP English for the first time and one request for guidance on better preparation for moderation. Five respondents expressed a need for guidance about the integration of APP into whole school practice.

> 'Advice on using with all pupils.'

> 'Advice on integrating outside pilot.'

> 'For school to have a long-term strategy in place for APP.'

13. Discussion

This section draws on evidence presented in the main body of the report to attempt to answer the key questions for the evaluation.

- Does APP provide an effective model for teacher assessment within key stage 2?

- Is the APP assessment model manageable?

It also identifies some factors which have impacted on the effectiveness and manageability of APP within the pilot.

Effectiveness

Figure 13.1 shows the cyclical integration of planning, teaching and assessment in APP. To be effective, all aspects of the cycle should work coherently. In addition, the outcomes from the assessment should be perceived as consistent, reliable and valuable by those who use them.
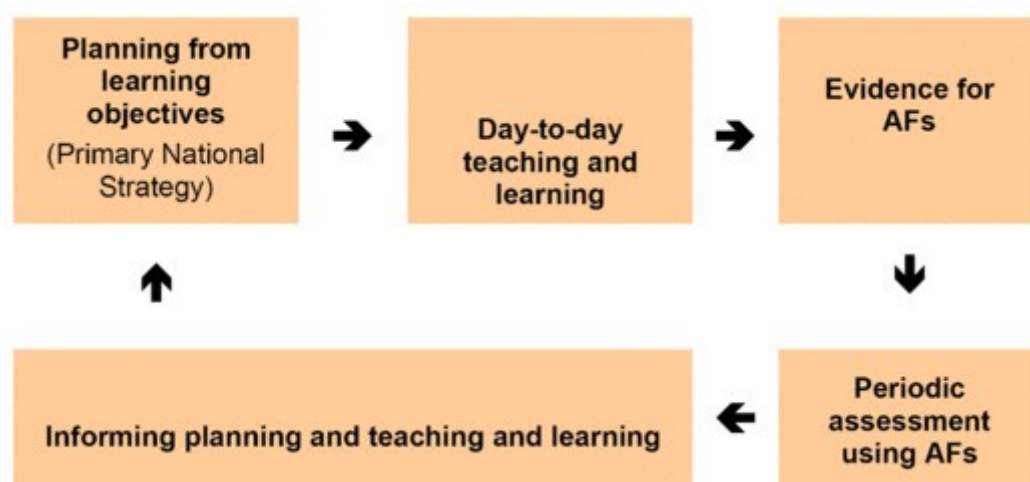
Figure 13.1

The available evidence from the pilot suggests that periodic assessment using APP has the potential to deliver effective teacher assessment. From the outset, teachers and headteachers have been convinced that using APP gives them a sound understanding of level criteria and supports them in making accurate assessments, adding to or deepening their previous knowledge of how their pupils are learning in reading, writing and mathematics. Completed assessment guidelines give a highly visual picture of strengths and weaknesses in a pupil's learning and this insight into the 'gaps' has been seen as a powerful means of informing planning so that specific learning needs can be addressed.

Two key changes in teacher views have been seen during the course of the pilot. The first is a move from a view which might be summarised as 'so-and-so is a level 2 child'. The use of APP has shown teachers that two children may well both be working securely within the same level but in fact have

very different patterns of strengths and weaknesses. This gives a clear message that different approaches may be needed to help these two children progress and is in line with a move towards more personalised learning.

The second change is a shift towards understanding how much more informative assessment can be when it draws on a range of opportunities for pupils to show what they can do, rather than on individual pieces of work or a 'one-off' structured assessment, including tests. Teachers have commented on the way in which APP has helped them to 'lift the ceiling' on their expectations because more open-ended questioning and activities and a range of work has pushed the boundaries of what children can achieve. Equally, the broader range of assessment activities can often reveal serious gaps in understanding or misconceptions which are not always evident in situations where credit is simply given for getting a 'correct' answer.

APP assessments not only identify gaps in the learning of individual pupils, they can also reveal lack of curriculum coverage in areas of reading, writing and mathematics. Teachers reacted very obviously to these insights in the pilot and have consistently commented on the strength of APP in informing planning. The introduction of the new framework at the end of the first year of the pilot has been reported by teachers as having made relationships between learning objectives and assessment focuses clearer and they appear to promote better integration of APP with effective planning.

There are of course two aspects to the relationship between APP assessment and planning. One is the way in which assessment outcomes can be used to inform and modify teachers' plans to address learning needs. The other is that teachers can use their plans to identify the possible future opportunities to pick up evidence to inform the next round of periodic reviews of achievement. Both need to be in place for really effective integration of planning and assessment.

In the pilot, the latter aspect (looking forward to see where assessment opportunities lie) tended to receive more attention in the earliest rounds of assessment as teachers needed to ensure that they would be able to make their next assessments. Outcomes could then be fed into the cycle for subsequent assessment rounds.

Teachers in the pilot believed that they had developed their own skills in what makes a 'good' assessment activity. They reflected on how limiting closed questions could be. A critical concept in effective APP assessment is the importance of 'distance' between the actual learning of a specific skill, method or technique and the opportunity to apply it, in order that the assessment can check that the teaching point has been assimilated by the pupil. During the course of the pilot, all sources suggested that there was a move away from the use of heavily scaffolded work or 'test-type' questions as evidence for APP assessment. This indicated that pupils had been given some opportunities to choose their own methods, approaches and writing form.

As mentioned above, assessment outcomes have often shown gaps in curricular provision which would obviously have major implications for planning. Unlike the 'gaps' for individual pupils, these cannot always immediately be addressed by class teachers. For example, a realisation that pupils have very little opportunity to access specific writing genres may reflect the types of texts available in school. In mathematics, pupils may simply not be getting the opportunities to collect primary data and decide how to represent it. More fundamental issues like this often needed action at subject coordinator or even school level to make the necessary improvements to curricular provision. Within the pilot, teachers recognised what they needed to improve but they could not always achieve the improvements individually. However, most were convinced that the experience of APP overall had a positive impact on their teaching.

To be an effective model of assessment, others apart from the teachers who 'own' the assessments must also be confident in the accuracy and worth of the outcomes. APP is based on nationally agreed standards exemplified within the standards files. The pilot has promoted accuracy and consistency through an expectation of in-school standardisation, in-school moderation and the involvement of individual moderators in the first year and the local authority-led moderation activities in the second year. Teacher and headteacher views suggest that they valued the inclusion of moderation and considered that it promoted objectivity in judgements and ensured that consistent national standards were being applied.

In the first year there appeared to be a lack of understanding in some cases of the role and importance of within-school processes to assure the quality and consistency of judgements. Evidence from questionnaire and moderator feedback showed that in the second year the status of these activities increased in those schools where APP was becoming embedded across more year groups and becoming part of established assessment practice. This was particularly so in schools where local authorities were organising regular external moderation meetings since there was a clear requirement that teachers would bring along collections which had already been subject to in-school moderation. As schools began to see a role for APP as a whole school assessment tool, they attached more importance to 'getting it right'.

At the moderation meetings in the second year, teachers were able to challenge and probe their own APP judgements and those of other teachers. Although it was sometimes rather uncomfortable for teachers to face challenges from their peers or to question the decisions of others, they considered the process of working together in this way highly beneficial. It was obvious that the APP criteria gave them a common language in which to describe achievement and being able to refer to objective, nationally agreed criteria gave them confidence in the value of their own and other teachers' judgements. Where local authorities took the opportunity to encourage teachers to take the lead in moderation discussions, this worked effectively and reinforced the message that teachers have a responsibility to ensure and promote standards in assessment.

Headteachers have referred to the robustness and reliability of APP assessment in key stage 2 and many obviously saw this form of periodic assessment as a potential replacement for existing systems. There was evidence across the two-year pilot of schools reducing their reliance on the use of national curriculum optional tests at the end of years 3, 4 and 5 as the main means of checking progress. Teachers recognised that APP can give important insights into achievement at any time during the course of a year and, most importantly, at a stage when action can be taken on the outcomes.

> 'Better to use teacher assessment while we can do something about what we find out rather than use tests and not look at them.'

For other schools, APP was seen as a useful assessment tool that could work alongside others and had its own particular contribution to make. Even in schools which intended to continue the use of optional tests, APP was seen as adding another dimension because of its ability not just to correctly define the 'level' of achievement, but also to show the profile of achievement within it. In a small subset of the pilot schools, APP assessment outcomes were used to inform the decisions for single level test entry. This report cannot offer any information on how effective this process was; there will be a separate full evaluation of the 'Making Good Progress' pilot.

APP therefore can complement the use of tests and teachers have been able to take account of outcomes from both types of assessment in their professional view of the progress made by their pupils.

By the end of the pilot, the confidence placed in APP outcomes was evident through their use in some schools' tracking and monitoring systems. Before APP, teachers' assessments were generally recorded at 'sub-level' (for example 3a and 2b) to feed into tracking systems. Whilst APP does not support sub-level judgements, the flowchart allows teachers to refine their level judgements as 'low', 'secure' or 'high' within a level. Teachers have generally equated these two approaches, regarding 'low' as broadly equivalent to sub-level 'c' and so on. In the analysis of progress conducted as part of the data collection for the evaluation, the performance of individual pupils was tracked from one assessment round to the next. This showed that using APP assessment outcomes to track progress gave patterns which were consistent with the accepted view that 'average' progress would be the equivalent of two 'sub-levels' per year. There is no apparent reason why APP teacher assessment should be incompatible with existing monitoring systems or why it could not be used for such purposes as target setting. Teachers have, however, identified that APP is more informative when looking at progress within a level and that it can be used effectively with parents or even pupils to show where progress has been made, even though there may have been no recent change in level outcome.

Manageability

The overwhelming message from the evaluation was that APP is not an 'easy' option. It requires the investment of considerable effort to achieve the benefits described above. Equally clear was the feedback that all the processes involved become easier and more manageable with practice. Teachers recognised that assimilation of the APP assessment focuses and criteria so that 'it's in your head' was key to becoming an efficient user of APP. Having a secure understanding of the assessment focuses meant that it was easier to identify:

- where the assessment opportunities are, in future planning

- evidence in pupils' work of their grasp of the relevant skills and concepts

- evidence of skills taught in English and mathematics being used in other areas of the curriculum.

Much less straightforward is the extent to which the process of 'gathering evidence' to inform APP judgements has been manageable and this is linked to concerns about how well the basic concept of a 'periodic assessment' has been understood.

The power of a periodic review is that it allows a teacher to reflect on achievement across the whole curriculum for reading, writing or mathematics to get a clear picture of where a pupil is in terms of their learning now and what the next steps are for progression. It was certainly the case that some teachers did not grasp this idea in the first year of the pilot. One of the problems seems to have been about the perceived purpose of the assessment guidelines. In many cases, teachers have, quite sensibly, made notes on the assessment guidelines of occasions when an activity or a comment or an observation has given them a 'piece of evidence' which they can then draw on when they come to carry out the next periodic review. However, this practice was translated by others into a kind of 'cumulative tick-list' with teachers constantly updating the guidelines as soon as they considered that the next criteria had been achieved and with the reflective element of periodic review reduced to a swift glance at the flowchart. In this scenario, the review process may be seen as quite manageable but the constant evidence gathering was not and it is certainly less likely to be effective.

In other cases, teachers were seen to go to great lengths to set up assessment activities which would allow them to acquire evidence against all the criteria, rather than drawing on what already existed in their everyday classroom interaction with pupils. This may have enabled them to make sound judgements but the process would neither be easy to manage nor would it be desirable, since assessment might be seen to be dominating the curriculum. There were particular problems with the interpretation of the mathematics assessment guidelines where the inclusion of several examples of ways in which pupils might demonstrate achievement of criteria led some teachers to feel the need to replicate every example in their evidence.

Perhaps the most problematic of the issues around the use of the assessment guidelines was that, faced with a lack of evidence, some teachers interpreted this not as an indication that they needed to look again at their coverage of the curriculum, but as a signal that they should teach the missing assessment focuses.

In the second year of the pilot, local authority meetings helped to clarify the nature of the process and, with greater experience, more teachers seemed to have a clear view of the nature and purpose of periodic review. Where teachers have clearly understood the difference between identifying evidence and reviewing the evidence available to inform their judgements, they have developed a range of strategies to capture their evidence, including the use of 'post-it' notes on planning, simple adjustments to marking techniques, photos and, wherever it has been possible, the involvement of teaching assistants to act as partners in observing pupils. All of these approaches help to make APP more manageable.

Taking part in the local authority-led activities during the second year certainly seemed to help teachers to improve their understanding of the principles of APP but there have been many concerns about what teachers have felt to be 'mixed messages' on what is or is not acceptable evidence. In the first year, mathematics moderators visiting schools encouraged teachers to draw on what they had seen or knew that pupils could do and not to rely too heavily on written work as the main source of evidence. In one-to-one discussions, the moderators could probe this verbal evidence and assure themselves of the accuracy of the judgements based on it. The message was consistent with that given to teachers at the start of the pilot about placing value on teachers' professional expertise. In contrast, some of the moderators for English were more likely to divide time on moderation visits between a review of a 'collection' of evidence away from the teacher with some follow up discussion and they encouraged teachers to prepare these collections by adding annotations, contextual information and commentaries. Some teachers, towards the end of the first year, appeared to equate the effort required for APP almost solely with the amount of time spent photocopying and organising their moderation 'portfolios'.

Within the pilot, particularly in the second year, there has been a considerable focus on external moderation. There is no doubt that the effort invested in moderation processes has promoted greater accuracy in judgements and also prompted teachers to improve the nature and range of evidence on which their judgements are based. The frequency of external moderation meetings within the pilot has been far greater than would be achievable or desirable once APP is established beyond the pilot and it was perhaps inevitable that in this second year the development work on models of moderation has led to an over-emphasis of this aspect of APP.

It is not, however, the time spent in meetings that teachers have found unmanageable, but the preparation for them in terms of photocopying evidence, annotating work and the completion of forms which record outcomes from in-school moderation and a review of evidence form. All this activity can be very important in ensuring that moderation is completed effectively and with the

greatest possible confidence that the outcomes from the moderation are accurate, whether or not the teacher is present during the moderation of their pupils' work. However, it cannot be replicated for every pupil for whom APP assessments are carried out and teachers have become confused about what is needed to support their own classroom assessments and what is needed for moderation purposes. The emphasis from moderators on the importance of well-presented, annotated 'collections' to support effective moderation has definitely had a negative impact on teacher perceptions of the manageability of the moderation process and, by extension, on their views on the manageability of APP generally.

There is concern among some teachers about the manageability of using APP beyond the small focus groups of pupils recommended in the pilot. By the end of the second year, almost all teachers seemed to feel that they could manage detailed periodic assessment for their focus group pupils but a substantial minority was worried that in the longer term this focus group approach would not be sufficient to underpin their assessment judgements for all pupils in their class. Indeed, some thought that this would deny the potential benefits for personalised learning to pupils outside the focus group and would therefore be inequitable. These teachers included some who felt that the use of APP for all pupils would be completely unmanageable and for them this is a serious barrier to the continued use of APP. Other teachers take the view that carrying out APP assessments for all pupils can be managed and is worth the effort involved. There may be a relationship here with their previous experience of assessment and tracking systems. In some schools, teachers are familiar with having to complete detailed recording sheets based on learning objectives on a regular basis for all pupils, so there have often been comments that using APP for all pupils would be no more difficult but potentially much more valuable. Of course, if teachers are being asked to continue to use these detailed tracking procedures alongside undertaking APP assessments, their perception of manageability will be very different to situations where APP is the main means of teacher assessment. Through the whole of the pilot, there were a small number of schools where APP was 'just something else for us to do', with the headteacher and senior leaders making no attempt to reduce effort expended on other approaches to assessment.

To be fully effective, APP has to be managed efficiently both within the classroom and at whole-school level. There was evidence that many of the pilot schools had moved forward rapidly to a whole school approach, seen in the integration of APP activities into staff meetings, planned cross-school or cross-phase moderations, agreements to reduce or abandon other forms of assessment when appropriate, and the development of consistent expectations about how and when APP assessments should be made and how these related to such factors as planning and target setting. All of the pilot schools started by implementing APP in a small way (in single subjects or one or two year groups) and in most cases have 'grown' their use of APP quite quickly. Teacher comments suggest that they think this is extremely important in ensuring that the process is not seen as too overwhelming and that it allows the perceived benefits to temper any early concerns about manageability.

Allowing time for teachers to develop and apply their skills in assessment was considered more important in ensuring the successful introduction of APP than all other factors by the pilot teachers. Within the first year of the project, this was supported in a small way through the provision of supply cover from central funds, sometimes added to by individual local authorities. Training and moderation was resourced centrally. In the second year, pilot schools continued to receive a small contribution to supply cover but it was clear from teacher feedback that this time was just a 'token' towards the effort required to really exploit the power and potential of APP. The challenge beyond the project will be to allow more schools to experience the benefits of APP without the support and resources offered in the pilot. Any plans for wider adoption of APP must not underestimate the resources required to support implementation, at least while teachers are developing their own skills and understanding of teacher assessment.

Appendix 1: Questionnaires

progi)8

**MONITORING CHILDREN'S PROGRESS
PILOT PROJECT 2006-2007**

Qualifications and
Curriculum Authority

First evaluation questionnaire

We would really welcome your views on your experience of the MCP project so far. Please take a few minutes to complete this questionnaire and return it in the envelope provided by 31 January 2007. Thank you.

Section 1: General information

(Please circle one response as appropriate in questions 1-3)

1. Which is your focus subject?   English   Mathematics
                                       52%   **48%**

2. Is your school using MCP for: English   Mathematics   Both
                                        29%   **25%**   **46%**

3. Is your school: First   Junior   Middle   Primary
                         17%   **21%**   **1%**   **61%**

4. Approximately how long have you been teaching? | **Av 9** | years

Section 2: Completing the assessment guidelines

5. How many pupils did you complete assessment guidelines for this term? | **Av 6-7** | pupils

6. About how long did it take to complete the guidelines for **each** pupil? | **Av 64** | mins

7. Overall, would you describe the process as:

    (Please circle one response only)

    very difficult   reasonably difficult   reasonably easy   very easy
      5%   **71%**   **24%**   **0%**

8. Which **one** aspect of making a level judgement did you find the most challenging?

    (Please tick one box only)

    finding evidence in children's work for the range of AFs | **57%%** |

    deciding a level for the individual AFs | **9%** |

    making a level judgement across a complete attainment target | **34%** |

9. Which assessment focus (AF) was the most difficult to make a judgement for in

    a. English:

    (Write the number of the AF in the box)

    reading | AF 7 (23) | writing | AF 1 (10) |

    OR

    b. mathematics

    (Write the name of the most challenging AF in each attainment target in the appropriate box e.g. if Reasoning was the most difficult to assess in the Handling Data attainment target, write 'Reasoning' in the Ma4 box)

    Ma1 | **Reasoning (32)** | Ma2 | **Solving numerical problems (11)** |

    Ma3 | **Position and movement (26)** | Ma4 | **Processing (16)** |

10. When you completed your assessment guidelines, how confident were you that the level judgements you had made were accurate:

    *(Please circle **one** response only)*

very confident	reasonably confident	not very confident	not at all confident
3%	**81%**	**14%**	**0%**

Section 3: Learning from the assessment outcomes

11. Which of the following **best** describes your view of the MCP assessment outcomes for your sample pupils:

    *(Please tick **one** box only)*

They did not tell me anything new about the sample pupils	**2%**
Completing the MCP assessment guidelines generally confirmed what I already thought the pupils could do	**32%**
Carrying out the MCP assessments added to my knowledge of what pupils could do	**34%**
The outcomes gave me important new insights into strengths and weaknesses for the sample pupils	**33%**

12. Has being part of the MCP project improved your understanding of what pupils need to do to achieve a particular NC level? *(Please tick)*

Yes | 92% | No | 8% |

If Yes, which one of the following would you say has been most helpful in improving your understanding?

*(Please tick **one** box only)*

attending the central training sessions | 6% |

standardisation activities in school | 6% |

completing the assessment guidelines | 63% |

moderating judgements with colleagues in school | 19% |

visit(s) from the external moderator | 6% |

13. Having completed the assessment guidelines, did you feel that the process gave you useful information about learning in your class as a whole? *(Please tick)*

Yes | 87% | No | 13% |

If Yes, can you give brief details of what you found out?

Are you intending to make any changes to teaching plans as a result?

Section 4: Processes in the MCP project

14. Which of the following project activities have you been involved in to date?

*(Tick **any** that apply)*

introducing MCP assessment process to other
teachers in your school

35%

in-school standardisation
*(making sure that all teachers involved have
the same understanding of NC level performance)*

87%

in-school moderation
*(checking that the assessment judgement are consistent
after the guidelines have been completed)*

79%

standardisation or moderation activities with other
schools apart from the central training
(these may have been organised by your LA)

13%

15. Was the information provided about completion and submission of the data collection forms clear and easy to follow?

Yes **80%**   No **19%**

16. How would you describe the visit(s) from your moderator in relation to supporting your work in the project?

*(Please circle **one** response only)*

Very valuable	Quite valuable	Not particularly valuable	Not at all valuable
35%	**45%**	**6%**	**1%**

17. Which of the following **best** describes the involvement of your headteacher in the MCP project?

*(Please tick **one** response only)*

Closely involved at
all stages so far **9%**

Takes an interest
and supports my/our
work **48%**

Supportive but
expects me/us to get
on with the work
independently **40%**

**MONITORING CHILDREN'S PROGRESS
PILOT PROJECT 2006-2007**

Second evaluation questionnaire

Thank you very much to those who returned the previous questionnaire. It is really important to the project to have your views. Please take a few minutes to complete this questionnaire and return it in the envelope provided by 30 March 2007. Thank you.

Section 1: General information

(Please circle one response as appropriate in questions 1-4)

1.	Which is your focus subject?	English **53%%**	Mathematics **47%**

2.	Is your school using MCP for:	English **28%**	Mathematics **25%**	Both **47%**

3.	Which year group do you teach?	Y3 **31**	Y4 **49**	Y5 **23**	Y6 **5**

4.	Are your MCP sample pupils taught in sets for the focus subject?	Yes **36%**	No **62%**

Section 2: Completing the assessments

5. Did you take part in an in-school standardisation before completing your assessments for this round? Yes **74%**   No **22%**

6. About how long did it take to complete the guidelines for **each** pupil? Av 45 mins

7. Did you find the assessment process easier than in the last assessment round?

Yes **80%**   No **20%**

If Yes, please can you say what made it easier this time?

8. Are you using the assessment guidelines for any pupils outside the data collection sample?

Yes for **18** pupils **11 seemed to be using for whole class**   No

9. Which assessment focus (AF) was the most difficult to make a judgement for in

a. English:

(Write the number of the AF in the box)

reading **AF 7 (26)**
AF 6 (8)
AF 5 (9)

writing **AF 2 (10)**
AF 4 (9)
AF 1 (5)

OR

b. mathematics

(Write the name of the most challenging AF in each attainment target in the appropriate box e.g. if

Representing was the most difficult to assess in the Handling Data attainment target, write

'Representing' in the Ma4 box)

Ma1 **Reasoning (27)**   Ma2 **Problem solving (8)**
Mental methods (8)

Ma3 **Position and movement (15)**   Ma4 **Processing (13)**

Can you explain what the problem was for the AF you have selected?

10. When you completed your assessment guidelines, how confident were you that the level judgements you had made were accurate:

*(Please circle **one** response only)*

very confident   reasonably confident   not very confident   not at all confident
14%   **78%**   **4%**   **0%**

Section 3: Planning and assessment

11. Did you adapt your teaching plans after you completed the last round of assessments?

Yes **76%**   No **20%**

If Yes, please give brief details of the changes

12.   Please look at the following statements and say whether you agree or disagree with each one by circling your chosen response.

I can look at my plans and see where there will be opportunities to assess specific AFs	Agree **91%**	Disagree **8%**
I need to use a test as well as ongoing classroom assessments to really know what my focus pupils can do	Agree **45%**	Disagree **50%**
I now organise lessons to teach specific assessment focuses	Agree **53%**	Disagree **42%**
Using MCP has prompted me to widen the range of activities I use for assessment	Agree **83%**	Disagree **15%**
MCP has made no difference to my teaching of the focus subject	Agree **12%**	Disagree **86%**
I would find it easy to explain the difference between a teaching objective and an assessment focus to a colleague	Agree **64%**	Disagree **31%**
MCP cannot give me a view of progress in my whole class	Agree **34%**	Disagree **59%**

13.   Have you made any changes to involve your pupils in their own assessments (e.g. peer assessment, through the use of traffic light indicators, review of personal targets etc)?

Yes **45%**   No **29%**

If Yes, please give brief details:

14.   Have you drawn on evidence from other subject areas?   Yes **71%**   No **29%**

If Yes, please say which:

15.   Have you used the MCP assessment outcomes to discuss progress with parents?

Yes **23%**   No **75%**

If Yes, what were the benefits or drawbacks of doing so?

Section 4: Working with others

*(Please circle **one** response only for questions 16-18)*

16. To what extent were you able to act on advice given by your external moderator about your assessment practice?

no action taken	made some changes	followed advice	not applicable as no advice given
2%	34%	56%	3%

17. How would you describe the in-school moderation process in this round compared to the first round of assessment?

more useful	as useful	less useful
17%	70%	4%

18. How confident are you that you and your colleagues have a consistent view of what pupils should be able to do at the national curriculum levels you have considered?

very confident	quite confident	not particularly confident	not at all confident
21%	72%	3%	2%

19. How many teachers in your school, other than those who came to the central training, are involved in using MCP now ?

   Av +1 Teachers

20. Are there plans to extend MCP into other year groups or subjects? Yes **42%**   No **26%**
    No decision 28%

    If Yes, is this expected to happen:

next term	next year	no decision made yet
7%	21%	31%

21. Which of the following has your headteacher/deputy headteacher been involved in:

   Tick any that apply

   in-school standardisation **55%**

   review of assessment outcomes from the previous assessment rounds
   (e.g. through looking at your completed guidelines or sharing the assessment report) **46%**

   in-school moderation **53%**

   staff meetings/training on extension of MCP **52%**

22. Please give brief details of any activities you have taken part in with other MCP schools or your local authority.

QCA Qualifications and
Curriculum Authority

**MONITORING CHILDREN'S PROGRESS
PILOT PROJECT 2006-2007**

Third evaluation questionnaire

Thank you very much to those who returned the previous questionnaire. It is really important to the project to have your views. Whether or not you have returned a questionnaire in previous rounds, please take a few minutes to complete this questionnaire and return it in the envelope provided by 12 July 2007. Completing the questionnaire will make sure that your views are reflected in the reports of the project. Thank you.

Section 1: General information

*(Please circle **one** response for each of questions 1-4)*

1. Which is your focus subject?

English **44%**   Mathematics **56%**

2. Is your school using MCP for:

English **20%**   Mathematics **30%**   Both **50%**

3. Which year group do you teach?

Y3 **28**   Y4 **43**   Y5 **31**   Y6 **9**

4. Are your MCP sample pupils taught in sets for the focus subject?

Yes **37%**   No **63%**

Section 2: MCP assessment processes in this assessment round

5. Did you take part in an in-school standardisation before completing your assessments for this round?

Yes **65%**   No **35%**

6. About how long did it take to complete the guidelines for an individual pupil? | **Av 40** | mins

7. Please give brief details of how you keep a record, for your own use, of the evidence that supports your assessment judgements.

8. Did you take part in in-school moderation after completing your assessments for this round?

    Yes **75%**   No **25%**

9. About how long did you spend preparing material for the visit by your external moderator?

    | **70** | Mins

10. Did you use the assessment guidelines for any pupils outside the data collection sample in this round?

    Yes for
    20   **13 appeared to be using with whole class**   No

11. When you completed your assessment guidelines, how confident were you that the level judgements you had made were accurate:

    *(Please circle **one** response only)*

very confident	reasonably confident	not very confident	not at all confident
34%	**65%**	**1%**	**0%**

12. Did you make any adjustments to your judgements as a result of the comments from your external moderator?   Yes   No
    27%   **73%**

Section 3: Using MCP assessment outcomes

13. Did you adapt your teaching plans after you completed the last round of assessments?

    Yes **70%**   No **30%**

    If Yes, please give brief details:

14. Have you identified any specific changes as a result of being involved in MCP:

    a. in your teaching   Yes **83%**   No **17%**

    If Yes, please give brief details:

b. in what your pupils can do   Yes**60%**   No**40%**

Please give brief details:

15. How have you used the outcomes of MCP assessments in your own class this year?

16. What would you say are the advantages and disadvantages of periodic teacher assessment and one-off tests?

Assessment type	Advantages	Disadvantages
Periodic teacher assessment		
One-off test		

17. Have you used the MCP assessment outcomes from this round to discuss progress with parents?

         Yes**30%**   No**70%**

If Yes, how did this work?

18.   If a teacher is to make detailed MCP assessments for a sample of pupils to get a view of progress in their whole class/group, what would be the best way to select the sample?

19.   What would be the best time of year to carry out MCP assessments in order to feed into planning for teaching and learning?

Section 4: Thinking about the future

20.   Will you be teaching the same group of pupils next year?   Yes **20%**   No **80%**
If Yes, please go to question 23, if No please go to question 21

21.   Will you be using MCP teacher assessment with your new class/group?   Yes **84%**   No **16%**

22.   Are you planning to pass on the details of MCP assessments you have made for your pupils this year to their new teacher?   Yes **87%**   No **13%**

23.   As far as you know does your school have plans to extend the use of MCP beyond the subjects/pupils/teachers involved in the pilot this year?   Yes **81%**   No **14%**
If Yes, please give brief details:

24.   What support would you like to see from your local authority for MCP teacher assessment?

QCA
Qualifications and
Curriculum Authority

**MONITORING CHILDREN'S PROGRESS
PILOT PROJECT 2006-2007**

Questionnaire for headteachers

We would really welcome your views on the MCP pilot project this year, whether or not your school will continue to take part next year. Please take a few minutes to share your views and experience of MCP. Responses will be analysed anonymously and reported to the project steering group with other findings from the evaluation to inform the next stages of the work.

Section 1: Manageability

1.   Please consider the following pairs of statements. For each pair, place a tick beside the statement which, in your view, best describes the experience of your teachers working with MCP this year.

Professionally challenging, but worthwhile	96%	Benefits did not justify the time and effort required	4%
Each assessment round has been as demanding as the last	16%	The processes have become easier to manage over time	80%
MCP assessment processes do not fit well with existing classroom practice	9%	MCP assessments can be slotted in to everyday teaching and learning	82%
Preparation for external moderator visits has resulted in additional paperwork	45%	The work done by teachers in carrying out assessments has given them all the necessary evidence to share with external moderators	56%
It has been possible to make use of time, including PPA time, for teachers to carry out the MCP assessment processes	69%	It has been difficult to create time for teachers to undertake the MCP assessment processes	22%
Detailed assessments for a small group of pupils have informed understanding of progress within the whole class	85%	Assessment of the focus group of pupils is not sufficient to give a view of progress in the class as a whole	7%

2.   Have you (or another member of your senior team) taken part in any of the following MCP activities?
        (Please tick any that apply)

        in-school standardisation   51%

        in-school moderation   73%

        review of assessment outcomes with teachers   65%

        training for teachers not attending the QCA-run sessions last autumn   24%

        external moderation visits   58%

Section 2 Impact on teaching and learning

3.   In your view, has the use of MCP improved the quality of teacher assessment in your school?   Yes **93%**   No **5%**
        If Yes, can you describe how?

4. Have the outcomes of MCP assessments been used to:

    a. inform teachers' future planning? Yes **95%** No **4%**

    b. inform discussion of progress with parents Yes **64%** No **25%**

5. Is there any evidence that as a result of MCP implementation, pupils are experiencing:

    a. more opportunities for independent learning? Yes **53%** No **29%**

    b. a greater range of learning opportunities? Yes **60%** No **27%**

    c. closer involvement in their own assessment? Yes **69%** No **24%**

6. Do you have any evidence to indicate that individual pupils are making better progress as a result of their involvement with MCP? Yes **42%** No **35%**

    If Yes, please give brief details.

    ```



    ```

Section 3 MCP and assessment practice

7. Please give brief details of your existing assessment practice excluding MCP *(e.g. termly/twice yearly/annual teacher assessment judgements for all pupils by sub-level, use of optional tests, etc.)*.

    ```



    ```

8. How are the outcomes of your existing assessments used within school? *(e.g. for setting school performance targets, teacher performance management, identifying curriculum priorities etc)*

    ```



    ```

9. Could MCP assessments replace any of your existing assessments? Yes **76%** No **11%**

    Please give reasons for your answer:

    ```



    ```

ASSESSING PUPILS' PROGRESS IN KS2 PILOT PROJECT 2007-2008

Autumn questionnaire for teachers

We would really welcome your views on your experience of the APP project early in the year. Please take a few minutes to complete this questionnaire on behalf of your school and return it to the address at the end of the questionnaire in the pre-paid envelope. Thank you.

Section A: General information

1. Please complete the grid below to show how APP is being used in your school this year.

    Enter the number of teachers using APP in each box as appropriate.

	Year 3	Year 4	Year 5	Year 6
English				
Mathematics				

2. Are any individual teachers using APP for both English and mathematics?   Yes **50%**   No **40%**

3. Has your school already started to use the renewed Primary Frameworks for:

    Literacy **7%**   mathematics **14%**   both **74%**   *(Please tick any that apply)*

Section B: Organisation of APP

4. Have you received updated versions of any of the following APP materials through your local authority? *(Please tick any that apply)*

    Updated assessment guidelines — **57%**

    Updated handbook — **24%**

    Updated collection of standards files — **31%**

5. Please give brief details of how teachers new to APP have been introduced to the process/materials (e.g. LA training session, in-school training, one-to-one support etc)

6. In general, what is the size of the focus group of pupils for whom detailed APP assessments are being made across the classes involved?   **Av 8**   Number of pupils

7. How have the groups been selected?

8. Have the APP standards files been used in school standardisation this term?

    Yes **57%**   No **40%**

If Yes, how and by whom?

[ ]

9. Approximately when were the assessment judgments made this term? *(Please circle one)*

Just before half term **5%**   Just after half term **12%**   Towards the end of term **64%**   Still to be completed **17%**

10. Has there been an in-school moderation this term?   Yes **72%**   No **26%**

If Yes, how many teachers, when was it held, were senior leaders involved?

[ ]

Section C: Impact of APP

11. Have APP assessments replaced any other types of assessment used previously in school?

If Yes, can you give brief details of what has changed?

[ ]

12. Have you or your colleagues used APP assessments to inform parent/teacher consultations this term?   Yes **29%**   No **64%**

13. Is there an agreed process in school for using APP outcomes to inform planning?

Yes **40%**   No **60%**   If Yes, please give brief details:

[ ]

Section D: Taking APP forward

14. What has been the most challenging aspect of using APP this term?

[ ]

15. Is there additional support that you and colleagues would like to see from your senior leadership team or from the LA?

[ ]

16. Has there been any opportunity this term to meet/work with other schools on APP?

Yes **59%**   No **41%**

ASSESSING PUPILS' PROGRESS IN KS2 PILOT PROJECT 2007-2008

Final questionnaire for teachers

Dear Colleague

The 2-year APP pilot is now drawing to a close. It is really important that we get the views of as many teachers as possible at the end of this pilot as the findings will help to shape future policy on teacher assessment.

So even if you have not returned previous questionnaires, please take a few minutes to complete this questionnaire on behalf of your school and return it to the address at the end of the questionnaire **by 11 July 2008**. A pre-paid envelope is provided.

It is also very important that APP assessment data for the summer term is returned to NFER by 13 June. If your school is using optional NC tests with the APP focus group pupils, the results from these need to be returned to NFER by 14 July. Please contact Dave Hereward (01753 637352, d.hereward@nfer.ac.uk) if you have any questions about the submission of assessment data.

Thank you.

Section A: General information

1. Please complete the grid below to show how APP is being used in your school this year.

    Enter the number of **teachers** using APP in each box as appropriate.

	Year 3	Year 4	Year 5	Year 6
English				
Mathematics				

2. Are any individual teachers using APP for both English and mathematics? Yes **45%**   No **40%**

Section B: Organisation of APP

3. APP is a process for periodic review of pupils' progress based on evidence collected over time. How often do you and your colleagues review progress using APP to derive a national curriculum level?

    Half-termly **25%**   *Termly* **70%**   *Other (please specify)* **3%**

4. Have any of the following activities taken place in school during the past year? *(please tick any that apply)*

    In-school training on use of APP   **80%**

    Standardisation of teacher judgements using standards files or the school's own materials   **67%**

    In-school moderation of APP assessments   **94%**

5. If you have been involved in in-school moderation please briefly describe the process (i.e. who attended the meeting, whether it was held in school time, how long it took, how many collections of pupils' work were moderated)

6. Have the APP assessments been used to inform teachers' assessments which feed into your school's process for tracking progress?

    Yes **83%**   No **17%**

7. Have any other forms of assessment activity (e.g. optional NC tests, individual writing assessment pieces, non-national curriculum progress tests) been replaced by APP?

    Yes **36%**   No **59%**

    If Yes, please give brief details

8. At the start of the pilot manageability was a concern for many teachers. Now that you have been using APP for some time, please could you rate the following aspects of APP in terms of their manageability:

    (Please circle 1-5 as appropriate where 1 is very manageable and 5 is not at all manageable)

	1	2	3	4	5
Collecting and recording evidence over time	5%	27%	45%	20%	2%
Reviewing evidence to update progress and decide the overall NC level	2%	44%	39%	11%	2%
Organising/attending in-school meetings for standardisation/moderation	5%	27%	34%	25%	8%
Preparing collections of work for external moderation	0%	11%	36%	36%	14%
Participation in external moderation meetings/activities	3%	22%	38%	22%	9%
Using outcomes of APP assessments to inform planning	28%	47%	19%	3%	2%

Section C: Impact of APP

9. Feedback from teachers suggests that APP might have an impact in a range of ways. Please look at the following table of possible benefits and put a tick in the column that reflects your own experience of the effects of APP.

	Improved by APP	Not changed by APP	Not relevant
Ability to identify good assessment opportunities in everyday classroom activities	92%	6%	0%
Understanding of what characterises performance within each NC level	92%	5%	0%
Ability to identify gaps in pupils' learning	95%	3%	0%
Involvement of pupils in assessing their own progress and what they need to do next	45%	52%	0%
Use of a wider range of texts/learning experiences for reading	47%	30%	22%
Offering more open-ended tasks and activities for mathematics	53%	27%	19%
Allowing pupils more opportunity to exercise independence and choice in their work	47%	48%	2%

Are there other benefits or drawbacks in your experience?

10.   Have you or your colleagues used APP assessments to inform parent/teacher consultations this term?   Yes **47%**   No**53%**

If Yes, can you comment on what you did and how well it worked?

11.   Are you using the new frameworks for literacy and/or mathematics?   Yes**89%**   No**9%**

If Yes, can you explain how the APP assessments work alongside the frameworks?

Section D: Taking APP forward

12.   What are the plans for the use of APP in your school from next September?

13.   Is there additional support that you and colleagues would like to see from your senior leadership team or from the LA?

14.   If you were asked to give one piece of advice, a 'Top Tip', to a teacher starting to use APP from September what would it be?

Appendix 2: Assessment guidelines

Assessment guidelines for reading, writing and mathematics were published by QCA and the National Strategies in 2008 and are available at http://nationalstrategies.standards.dcsf.gov.uk/node/20683.

Appendix 3: Whole sample level distributions 2006/7

Reading
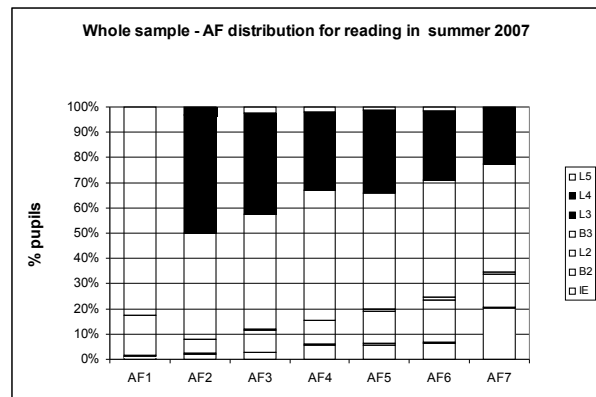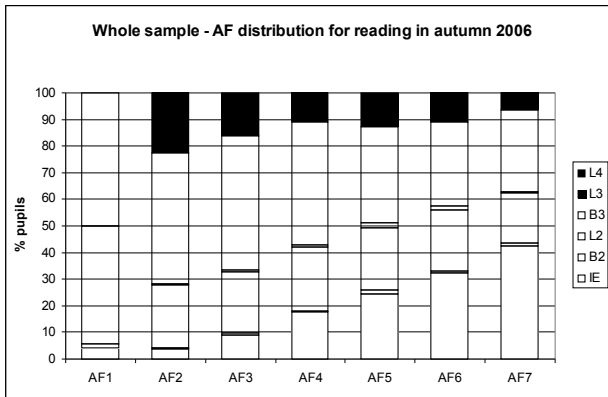
Writing
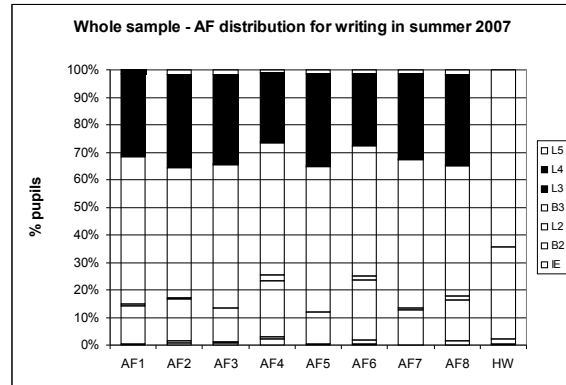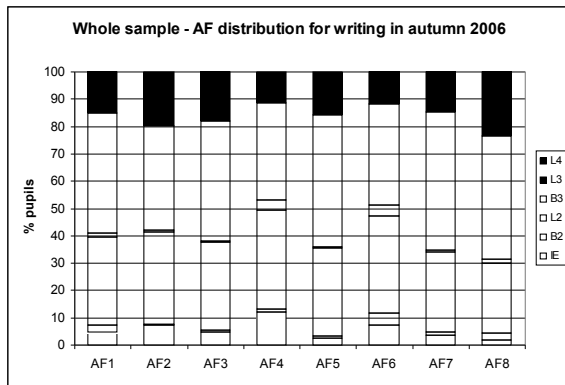
Ma1 – Using and applying mathematics
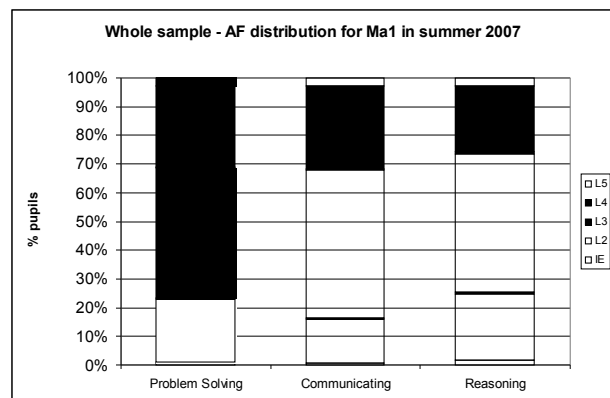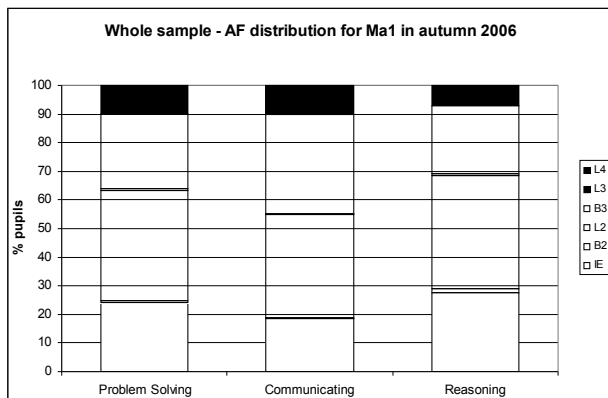
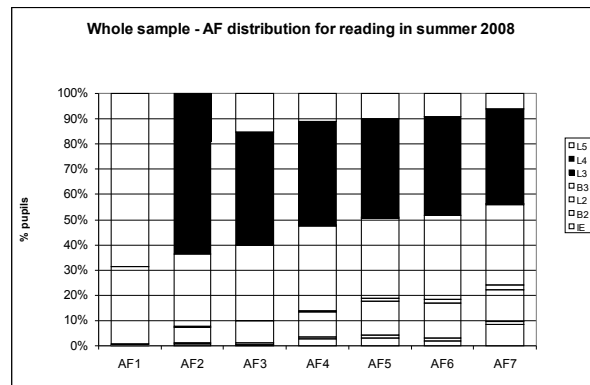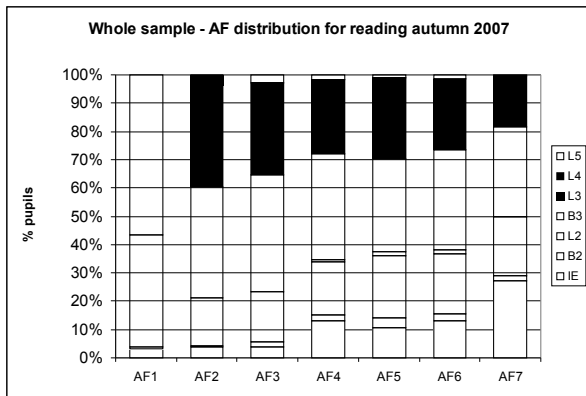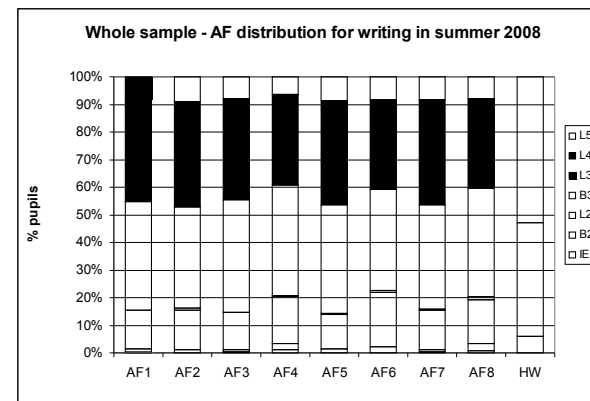## Ma2 – Numbers and the number system
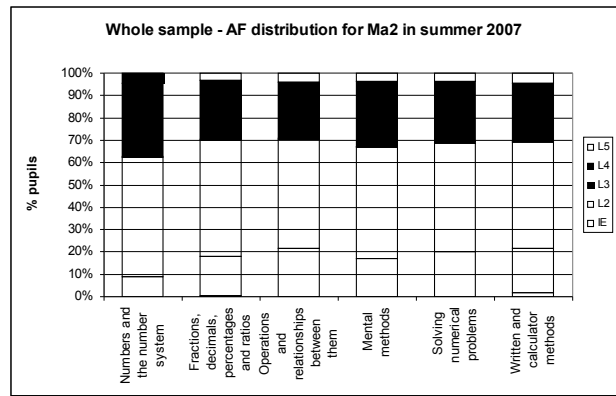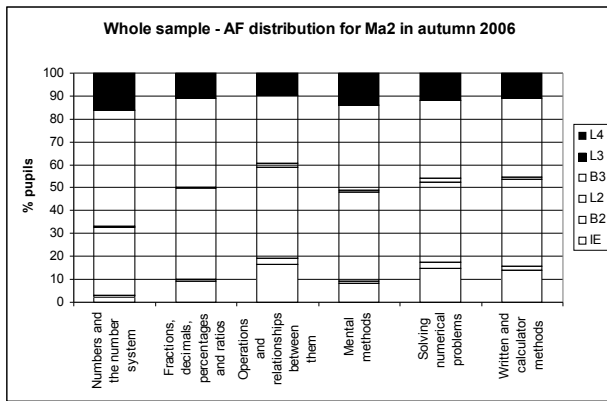
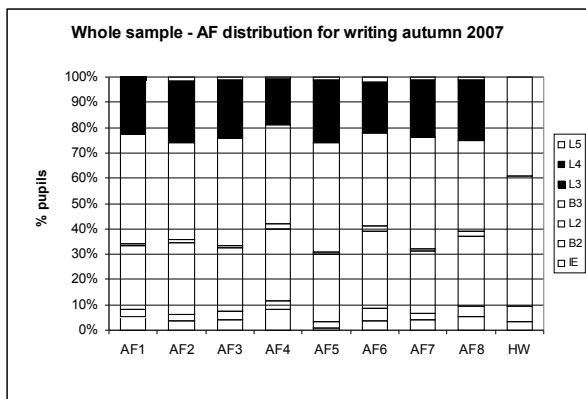## Ma3 – Shape, space and measures
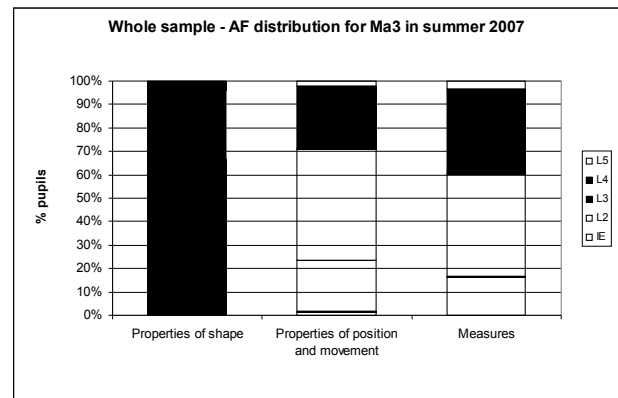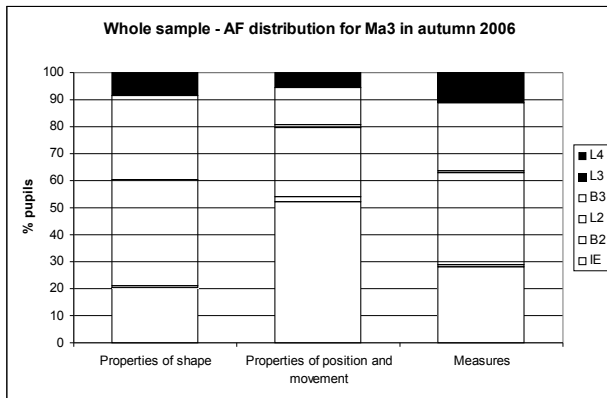
Ma4 – Data handling
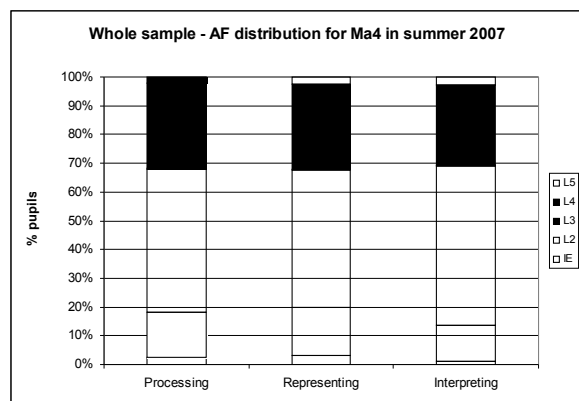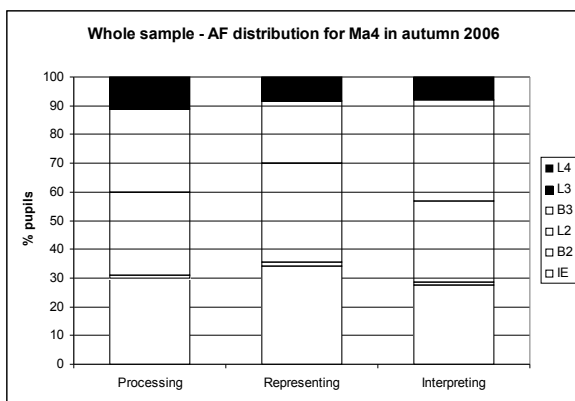
Appendix 4: Whole sample level distributions 2007/8

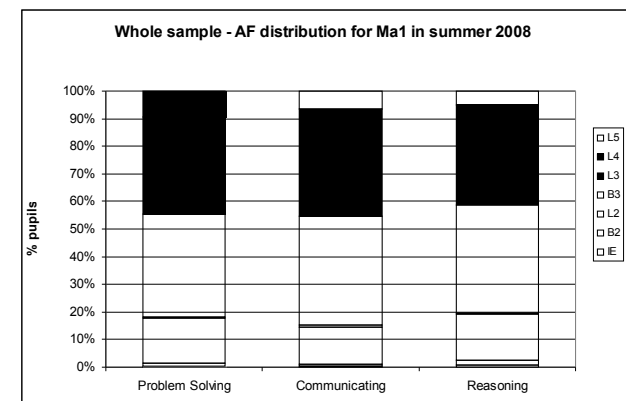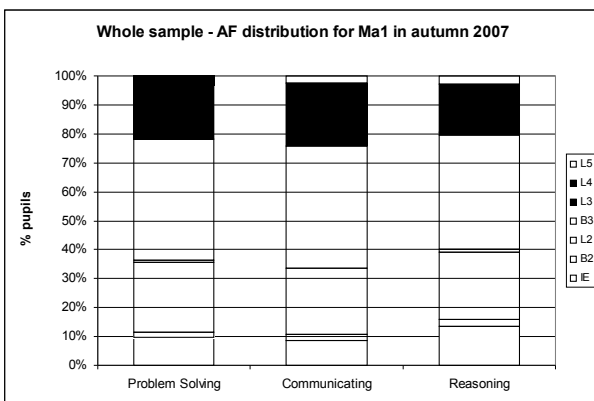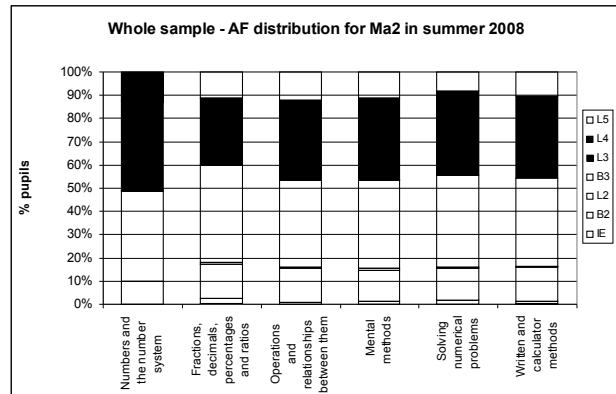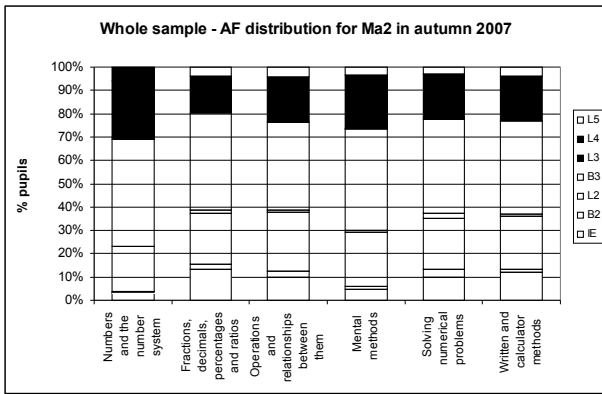## Reading

Writing
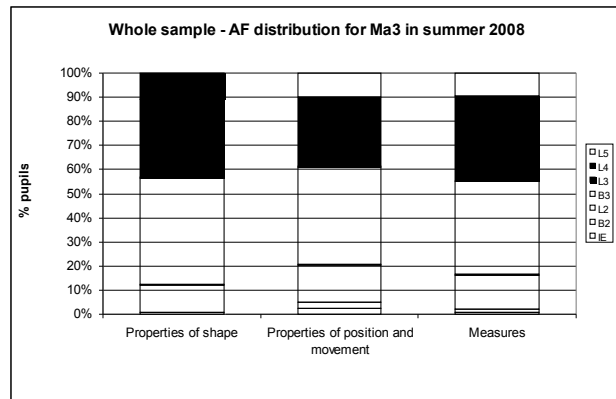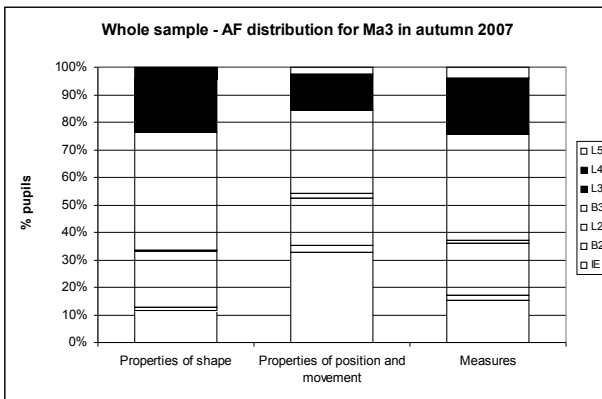

Ma1 – Using and applying mathematics

Ma2 – Numbers and the number system

Whole sample - AF distribution for Ma2 in autumn 2007

Whole sample - AF distribution for Ma2 in summer 2008

Ma3 – Shape, space and measures

Whole sample - AF distribution for Ma3 in autumn 2007

Whole sample - AF distribution for Ma3 in summer 2008

Ma4 – Data handling

Whole sample - AF distribution for Ma4 in autumn 2007

Whole sample - AF distribution for Ma4 in summer 2008

Appendix 5: Summary of pupil assessment outcomes 2006/7

English – Year 3

Year 3 Reading Overall Levels

Year 3 Writing Overall Levels

English – Year 4

Year 4 Reading Overall Levels

Year 4 Writing Overall Levels

English – Year 5

Year 5 Reading Overall Levels

Year 5 Writing Overall Levels

English – Year 6

Year 6 Reading Overall Levels

Year 6 Writing Overall Levels

Mathematics – Year 3

Year 3 Overall Mathematics Level

Year 3 Ma1 Overall Levels

Year 3 Ma2 Overall Levels

Year 3 Ma3 Overall Levels

Year 3 Ma4 Overall Levels

Mathematics – Year 4

Year 4 Overall Mathematics Level

Year 4 Ma1 Overall Levels

Year 4 Ma2 Overall Levels

Year 4 Ma3 Overall Levels

Year 4 Ma4 Overall Levels

Mathematics – Year 5

Year 5 Overall Mathematics Levels

Year 5 Ma1 Overall Levels

Year 5 Ma2 Overall Levels

Year 5 Ma3 Overall Levels

Year 5 Ma4 Overall Levels

Mathematics – Year 6

Year 6 Overall Mathematics Levels

Year 6 Ma1 Overall Levels

Year 6 Ma2 Overall Levels
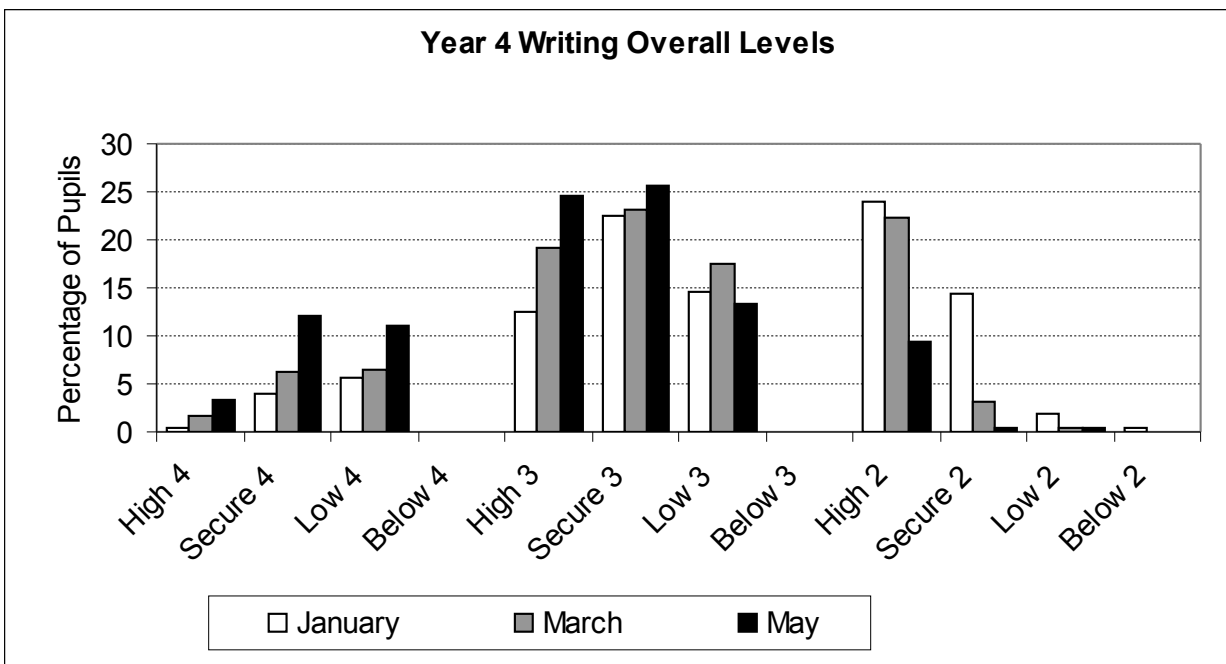

Year 6 Ma3 Overall Levels

Year 6 Ma4 Overall Levels

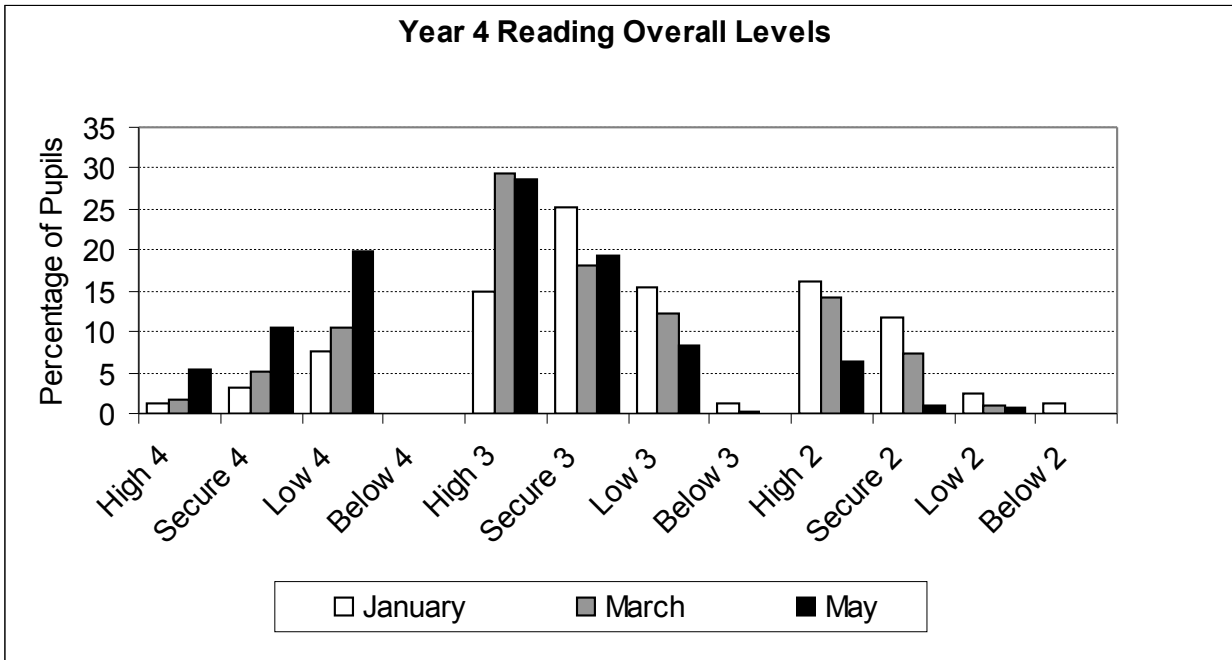# Appendix 6: Summary of pupil assessment outcomes 2007/8

English – Year 3

Year 3 Reading Overall Levels

Year 3 Writing Overall Levels

English – Year 4

Year 4 Reading Overall Levels

Year 4 Writing Overall Levels

English – Year 5

Year 5 Reading Overall Levels

Year 5 Writing Overall Levels

English – Year 6

Year 6 Reading Overall Levels

Year 6 Writing Overall Levels

Mathematics – Year 3

Year 3 Overall Mathematics Level

Year 3 Ma1 Overall Levels

Year 3 Ma2 Overall Levels

Year 3 Ma 3 Overall Levels

Year 3 Ma4 Overall Levels

Mathematics – Year 4

Year 4 Overall Mathematics Level

Year 4 Ma1 Overall Levels

Year 4 Ma2 Overall Levels

Year 4 Ma3 Overall Levels

Year 4 Ma4 Overall Levels

Mathematics – Year 5

Year 5 Overall Mathematics Level

Year 5 Ma1 Overall Levels

Year 5 Ma2 Overall Levels

Year 5 Ma3 Overall Levels

Year 5 Ma4 Overall Levels

Mathematics – Year 6

Year 6 Overall Mathematics Level

Year 6 Ma1 Overall Levels

Year 6 Ma2 Overall Levels

Year 6 Ma3 Overall Levels

Year 6 Ma4 Overall Levels

Appendix 7: Teacher judgements and moderator judgements 2006/7

Table 1 – Analysis of teacher-moderator differences in judgements for Ma1 in summer and spring (per cent moderated judgements)

Moderators' judgements	AF Problem solving	AF Communicating	AF Reasoning	Overall level for Ma1
Higher than school judgement	1(1)	2(5)	3(6)	5(4)
Same as school judgement	97(95)	97(94)	95(94)	90(93)
Lower than school judgement	2(3)	2(1)	3(0)	5(7)

Table 2 – Analysis of teacher moderator differences in judgements for Ma2 in summer and autumn assessment rounds (per cent moderated judgements)

Moderators' judgements	AF Numbers	AF Fractions	AF Operations	AF Mental methods	AF Solving numerical problems	AF Written & calculator methods	Overall level for Ma2
Higher than school judgement	2(9)	4(8)	3(8)	5(10)	2(8)	3(90	1(9)
Same as school judgement	98(89)	93(85)	96(89)	94(87)	98(89)	96(91)	98(83)
Lower than school judgement	1(2)	2(7)	1(3)	1(3)	1(3)	1(0)	1(8)

Table 3 – Analysis of teacher moderator differences in judgements for Ma3 summer and spring assessment rounds (per cent moderated judgements)

Moderators' judgements	AF Properties of Shape	AF Position and movement	AF Measures	Overall level for Ma3
Higher than school judgement	4(2)	2(5)	2(7)	4(7)
Same as school judgement	94(97)	96(94)	97(92)	93(89)
Lower than school judgement	2(1)	2(1)	1(1)	3(5)

Table 4 – Analysis of teacher moderator differences in judgements for Ma4 in summer and spring (per cent moderated judgements)

Moderators' judgements	AF Processing	AF Representing	AF Interpreting	Overall level for Ma4
Higher than school judgement	3(2)	6(5)	2(5)	4(9)
Same as school judgement	95(97)	92(93)	96(93)	92(85)
Lower than school judgement	2(1)	2(2)	2(2)	4(6)

Table 5 – Analysis of teacher moderator differences in judgements for reading in summer and spring (per cent moderated judgements)

Moderators' judgements	AF1	AF2	AF3	AF4	AF5	AF6	AF7	Overall
Higher than school	2(0)	1(3)	1(6)	1(2)	1(5)	1(6)	0(2)	3(2)
Same as school	98(100)	98(93)	94(88)	94(87)	94(85)	95(94)	97(94)	79(75)
Lower than school	0(0)	1(4)	5(6)	6(10)	6(11)	5(3)	3(3)	18(22)

[Overall figures from spring and summer term only]

Table 6 – Analysis of teacher moderator differences in judgements for writing in summer and spring (per cent moderated judgements)

Moderators' judgements	AF1	AF2	AF 3	AF 4	AF5	AF6	AF7	AF8	Overall
Higher than school	3(5)	0(0)	0(5)	1(7)	6(7)	2(2)	2(2)	0(2)	8(6)
Same as school	92(88)	93(93)	82(86)	93(81)	83(88)	90(86)	94(91)	99(98)	69(73)
Lower than school	4(7)	7(7)	18(8)	6(12)	10(6)	8(11)	3(8)	1(0)	23(21)

[Overall figures from spring and summer only]

Appendix 8: Comparison of school and moderation overall level judgements over three terms of the moderation pilot 2007/8

Table 3

Key stage/subject	Autumn		Spring		Summer	
	Number of collections moderated	Per cent overall level judgement confirmed	Number of collections moderated	Per cent overall level judgement confirmed	Number of collections moderated	Per cent overall level judgement confirmed
KS3 reading	10	60%	32	75%	-	
KS3 writing	29	69%	20	80%	-	
KS2 reading	8	63%	66	55%	68	66%
KS2 writing	57	84%	37	70%	15	67%
KS2 mathematics Ma1	5	60%%	60	73%	69	72%
KS2 mathematics Ma2	35	80%	37	62%	3	100%
KS2 mathematics Ma3	N/A	N/A	54	81%	4	25%
KS2 mathematics Ma4	N/A	N/A	28	89%	3	33%

Appendix 9: Advice from pilot teachers to teachers new to APP

Use post-it notes to annotate evidence as often as possible.
Choose children from across the range of ability in the class.
Have APP materials to hand at all times, explain process to children and use a different colour highlighter pen for each half term to show progress.
Use your assessments to inform your planning it is great for identifying next steps.
Regularly update your grids to help with manageability.
Collect data for one to two pupils at a time.
Focus on one AF at a time to become familiar with it.
Bear in mind that the process does become quicker and more efficient with time and experience.
Take it slowly – don't panic about collecting evidence.
Start small. It needs to be carefully planned in with clear milestones to monitor progress.
Take time to carefully read the criteria, and ensure you really understand them before actually assessing.
English takes more time to get your head around than the maths. Start with small groups and don't attempt to collect evidence for every single AF.
Know what specific assessment focus you are trying to gather evidence for.
Don't be put off by the paper. It gets easier and makes sense so don't give in too quickly. Use highlighters to mark attainment (change colour each term).
Buy lots of post-its! Very useful in jotting down comments and observations. Remember it doesn't have to be a written task – go for practical U&A opportunities.
Persevere!
Start small!
Don't panic! – collect post it notes – as they do in foundation stage.
Keep a copy of the grid on the wall, become as familiar as possible.
Use it as a tool, not as an additional element.
Select the child very carefully so there is one in each differentiated group as they are 'typical'.
Plan in more opportunities for independent tasks at regular intervals to use for assessment.
Ensure independent writing is planned regularly each week so children get used to trying out, applying writing skills – share your thoughts/issues with others.
DON'T PANIC – it really is useful!
Don't panic about trying to do everything all at once – choose one strand.
It will probably take a new teacher a while to adapt to the system. However once one has become familiar with the process it becomes a lot easier (don't panic!).
Regular recording of assessment so as to inform planning.
Don't panic – it's not as bad as it looks.

Be very aware of the AFs and plan carefully to cover them all.
Put the AFs in your room where you can see them all the time, it's quicker and easier once you know all the AFs.
Don't give up – once you are used to the process the benefits for pedagogy and practice are clear.
You need to embrace this process as an ethos not just completing paperwork – APP is all about clearly identifying the next steps in a child's learning against a benchmark of standardisation against NC levels.
Don't highlight until five to six instances of a particular bullet point have been identified.
Reciprocal reading gives super opportunities for collecting evidence of children's abilities in reading.
Be familiar with the AFs and write them on work/planning at time of using them for easy reference.
Be organised – organise assessment folders well.
Put assess guidelines sheet in planning folder to be used weekly rather than in assessment folder.
Don't try to use the guidelines for *every* pupil. Begin by using two children at three different levels.
Incorporate into planning and save examples/photos etc of work in other curriculum areas.
Use APP to identify next steps and inform your planning.
Moderate regularly and include within planning. Keep evidence.
Use guidelines as a benchmark to represent a group of similar ability – do *not* attempt to use guidelines for each child.
Choose assessment focus wisely to reflect what evidence you *need* to collect.
Learn the AFs.
Have all levels in front of you when you start & work from the bottom up. Initially concentrate on a couple of AFs and get familiar.
Funded time for moderating (in school moderation) training for teachers.
Work as a team to start with – you won't get your head around it otherwise. Start with two children only.
Don't panic! You are setting out on a journey and it takes time. Get really familiar with the assessment focuses and focus on a small group to start with. Trust your instincts; look carefully and listen to children. Be prepared to change planning to incorporate opportunities for assessment.
Plan assessment opportunities in regularly – know what you will assess.
Ensure relevant assessments are available at planning meetings to support teacher knowledge and understanding and for differentiation. Keep MAI highly focused.
Don't be overwhelmed by the amount of work – over time it becomes much more manageable.
Start small, but expand quickly.
It's probably easier to pick one piece of child's work and see how many AFs you can cross-reference it with – some pieces will show multiple skills.
Start with maths or English – not both at the same time.
Take it slowly to familiarise yourself with the AFs.
Start small – take either writing or reading or numeracy and make sure that you fully understand what the statements mean.
Start with just two AFs and work on those first, then familiarise with 2 more etc. DON'T RUSH IT!

Appendix 10: Headteacher summaries of existing assessment practice

Description	Fit with MCP?
Termly internal tests, annual use of optionals	No
Termly assessment for all, predicted targets set annually	Yes
Half-termly (AfL), end-of-year optionals	Yes
Regular use of ISP targets (learner ladders) and termly assessment	No
Sub-levels assessed 3 times per year in reading, 5 times in writing and use of optionals	Yes
Termly review of targets and key objectives assessments, yearly optional tests	Yes
AfL practice embedded – used for continuous teacher assessment	Yes
Optional tests, termly assessment tasks, termly monitoring of sub-levels	No response
Termly assessment, KS2 mock SATs, end of year 'non-reported' SAT	Yes
Termly TA judgement for all pupils by sub-level, use of optional tests	Yes
Half-termly assessment by sub-level, optional test in summer, reading recovery assessment half-termly	No
Termly TA by sub-level, end of year QCA	Yes
TA termly by sub-level with half termly review, optional SATs, phonics assessment, reading and writing baseline	No
Termly TA by sub-level, some use of optional tests during assessment work in September to inform planning	Yes
Target-setting autumn term, review spring and summer, optional SATs	No response
Half-termly TA for core subjects	No response
Twice-yearly QCA, termly/half-termly TAs	Yes
Initial termly baseline assessments (reading, spelling and NVR), end-of-term assessment - optional test and TA	Yes
Twice-yearly optional tests, children involved in self-assessment	Yes
Termly TA all pupils by sub-level, use of optional tests	Yes
Twice yearly all pupils by sub-levels, optional tests	Yes
Twice termly use of sub-levels, QCA SATs	Yes
Reading and writing half-termly, unit assessment in maths, TA for pupils by sub-levels and optionals	Yes
Termly TA for all pupils by sub-level, optional tests	Yes
Optional tests	Yes
Termly TA for sub-levels and comparison to optionals at end of year	Yes
Termly TA, all pupils by sub-levels, optionals	Yes
Termly sub-level review, optionals	Yes

Description	Fit with MCP?
Key objectives completed termly for maths, optionals 3 times a year	Yes
Termly sub-levels in maths and writing, twice yearly QCA tests	Not sure
Optionals, NFER tests, TA and termly tracking sheets	Yes
Annual diagnostic baseline, end –of-year optionals (not writing), whole school assessment activities termly for maths/science, unaided writing half-termly, pupil tracking half termly	Yes
Half-termly assessment, optional SATs	Yes
Termly tracking of TA, optional tests	Yes
Twice-yearly TA, optional tests	No
Twice-termly TA, tests as diagnostic	Yes
Termly TA in writing, optional test reading twice yearly, optional maths twice yearly, Salford and Vernon reading tests	Yes
Termly TA, QCA tests as diagnostic	No reply
Termly TA sub-levels, optional tests	Not sure
Termly test and TA for maths, science and English by sub-level, the summer term test is QCA optional	Yes
Optional tests in summer, termly sub-levelled TA for core and ICT	Yes
Twice-termly by sub-level against targets by TA, end of year optionals used as extra evidence	Yes
Whole school push on daily ongoing assessment, QCA tests once a year	Yes
Termly assessment with tracking	Yes
Termly formal assessment using QCA tests	Yes
Termly TA with optional test in summer term	Yes
Half-termly TA writing, optional tests	Yes
Termly assessment (key objectives, optional tests, writing samples, online assessments, end of unit assessments)	Yes
Half-termly writing samples, annual optional tests	Yes
Set targets reviewed termly, optional tests	Yes
Moderated writing samples, optional tests	Yes
Optional tests, individual teachers often test termly rather than carrying out a TA	Yes
Twice yearly optional tests, key objectives tick sheet	Yes
Termly TA by sub-level, termly writing assessment, optional tests	Yes

Appendix 11: Frequency analysis – APP assessment vs optional tests summer 2007

Reading – Whole sample by sub-levels

			Assessment type		Total
			MCP reading	Optional reading test	
Reading sublevel	Low 2	Count	3	1	4
		% within measure	0.6%	0.2%	0.4%
	Secure 2	Count	7	9	16
		% within measure	1.4%	1.8%	1.6%
	High 2	Count	33	17	50
		% within measure	6.8%	3.5%	5.1%
	Below 3	Count	0	7	7
		% within measure	0.0%	1.4%	0.7%
	Low 3	Count	44	28	72
		% within measure	9.0%	5.7%	7.4%
	Secure 3	Count	89	57	146
		% within measure	18.2%	11.7%	15.0%
	High 3	Count	127	67	194
		% within measure	26.0%	13.7%	19.9%
	Low 4	Count	87	69	156
		% within measure	17.8%	14.1%	16.0%
	Secure 4	Count	60	156	216
		% within measure	12.3%	32.0%	22.1%
	High 4	Count	36	41	77
		% within measure	7.4%	8.4%	7.9%
	Low 5	Count	2	0	2
		% within measure	0.4%	0.0%	0.2%
	Secure 5	Count	0	36	36
		% within measure	0.0%	7.4%	3.7%
Total		Count	488	488	976
		% within measure	100.0%	100.0%	100.0%

Reading by year group

				Assessment type		
				MCP reading	Optional reading test	Total
Year 3 reading level		2	Count	25	20	45
			% within measure	17.9%	14.3%	16.1%
		3	Count	107	71	178
			% within measure	76.4%	50.7%	63.6%
		4	Count	8	49	57
			% within measure	5.7%	35.0%	20.4%
	Total		Count	140	140	280
			% within measure	100.0%	100.0%	100.0%
Year 4 reading level		2	Count	14	7	21
			% within measure	7.2%	3.6%	5.4%
		3	Count	101	57	158
			% within measure	51.8%	29.2%	40.5%
		4	Count	80	131	211
			% within measure	41.0%	67.2%	54.1%
	Total		Count	195	195	390
			% within measure	100.0%	100.0%	100.0%
Year 5 reading level		2	Count	4	7	11
			% within measure	2.6%	4.6%	3.6%
		3	Count	52	24	76
			% within measure	34.0%	15.7%	24.8%
		4	Count	95	86	181
			% within measure	62.1%	56.2%	59.2%
		5	Count	2	36	38
			% within measure	1.3%	23.5%	12.4%
	Total		Count	153	153	306
			% within measure	100.0%	100.0%	100.0%

Writing – Whole sample by sub-levels

			Assessment type		Total
			MCP writing	Optional writing test	
Writing sublevel	Low 2	Count	3	4	7
		% within measure	0.6%	0.8%	0.7%
	Secure 2	Count	11	23	34
		% within measure	2.3%	4.8%	3.6%
	High 2	Count	42	49	91
		% within measure	8.8%	10.3%	9.5%
	Below 3	Count	1	1	2
		% within measure	0.2%	0.2%	0.2%
	Low 3	Count	69	65	134
		% within measure	14.5%	13.6%	14.0%
	Secure 3	Count	103	101	204
		% within measure	21.6%	21.2%	21.4%
	High 3	Count	115	68	183
		% within measure	24.1%	14.3%	19.2%
	Low 4	Count	59	36	95
		% within measure	12.4%	7.5%	10.0%
	Secure 4	Count	55	107	162
		% within measure	11.5%	22.4%	17.0%
	High 4	Count	17	11	28
		% within measure	3.6%	2.3%	2.9%
	Low 5	Count	2	0	2
		% within measure	0.4%	0.0%	0.2%
	Secure 5	Count	0	12	12
		% within measure	0.0%	2.5%	1.3%
Total		Count	477	477	954
		% within measure	100.0%	100.0%	100.0%

Writing by year group

				Assessment type		Total
				MCP writing	Optional writing test	
Year 3 writing level		2	Count	29	44	73
			% within measure	23.6%	35.8%	29.7%
		3	Count	90	68	158
			% within measure	73.2%	55.3%	64.2%
		4	Count	4	11	15
			% within measure	3.3%	8.9%	6.1%
	Total		Count	123	123	246
			% within measure	100.0%	100.0%	100.0%
Year 4 writing level		2	Count	19	32	51
			% within measure	9.1%	15.3%	12.2%
		3	Count	130	113	243
			% within measure	62.2%	54.1%	58.1%
		4	Count	60	64	124
			% within measure	28.7%	30.6%	29.7%
	Total		Count	209	209	418
			% within measure	100.0%	100.0%	100.0%
Year 5 writing level		2	Count	9	1	10
			% within measure	6.2%	0.7%	3.4%
		3	Count	67	53	120
			% within measure	46.2%	36.6%	41.4%
		4	Count	67	79	146
			% within measure	46.2%	54.5%	50.3%
		5	Count	2	12	14
			% within measure	1.4%	8.3%	4.8%
	Total		Count	145	145	290
			% within measure	100.0%	100.0%	100.0%

Mathematics – Whole sample by sub-levels

| | | | Assessment type | | |
			MCP mathematics	Optional mathematics test	Total
Mathematics sublevel	Low 2	Count	0	1	1
		%	0.0%	0.2%	0.1%
	Secure 2	Count	15	12	27
		%	3.4%	2.8%	3.1%
	High 2	Count	44	78	122
		%	10.1%	17.9%	14.0%
	Below 3	Count	4	4	8
		%	0.9%	0.9%	0.9%
	Low 3	Count	76	82	158
		%	17.4%	18.8%	18.1%
	Secure 3	Count	90	74	164
		%	20.6%	17.0%	18.8%
	High 3	Count	88	64	152
		%	20.2%	14.7%	17.4%
	Below 4	Count	1	0	1
		%	0.2%	0.0%	0.1%
	Low 4	Count	54	56	110
		%	12.4%	12.8%	12.6%
	Secure 4	Count	39	30	69
		%	8.9%	6.9%	7.9%
	High 4	Count	21	26	47
		%	4.8%	6.0%	5.4%
	Low 5	Count	4	5	9
		%	0.9%	1.1%	1.0%
	Secure 5	Count	0	4	4
		%	0.0%	0.9%	0.5%
Total		Count	436	436	872
		%	100.0%	100.0%	100.0%

Mathematics by year group

				Assessment typ		
				MCP mathematics	Optional mathematics test	Total
Year 3 mathematics level		2	Count	44	65	109
			% within measure	27.0%	39.9%	33.4%
		3	Count	114	85	199
			% within measure	69.9%	52.1%	61.0%
		4	Count	5	13	18
			% within measure	3.1%	8.0%	5.5%
	Total		Count	163	163	326
			% within measure	100.0%	100.0%	100.0%
Year 4 mathematics level		2	Count	13	26	39
			% within measure	9.8%	19.5%	14.7%
		3	Count	86	78	164
			% within measure	64.7%	58.6%	61.7%
		4	Count	34	29	63
			% within measure	25.6%	21.8%	23.7%
	Total		Count	133	133	266
			% within measure	100.0%	100.0%	100.0%
Year 5 mathematics level		2	Count	6	4	10
			% within measure	4.3%	2.9%	3.6%
		3	Count	55	57	112
			% within measure	39.3%	40.7%	40.0%
		4	Count	75	70	145
			% within measure	53.6%	50.0%	51.8%
		5	Count	4	9	13
			% within measure	2.9%	6.4%	4.6%
	Total		Count	140	140	280
			% within measure	100.0%	100.0%	100.0%

Appendix 12: Frequency analysis – APP assessment vs optional tests summer 2008

Reading – Whole sample by sub-levels

			Assessment type		Total
			APP	Optional/KS2 test	
Reading sublevel	High 5	Count	4	0	4
		% within measure	0.7%	0.0%	0.4%
	Secure 5	Count	36	171	207
		% within measure	6.3%	30.2%	18.2%
	Low 5	Count	56	0	56
		% within measure	9.8%	0.0%	4.9%
	High 4	Count	103	35	138
		% within measure	18.0%	6.2%	12.1%
	Secure 4	Count	97	179	276
		% within measure	16.9%	31.6%	24.2%
	Low 4	Count	72	35	107
		% within measure	12.6%	6.2%	9.4%
	High 3	Count	68	45	113
		% within measure	11.9%	7.9%	9.9%
	Secure 3	Count	60	43	103
		% within measure	10.5%	7.6%	9.0%
	Low 3	Count	38	27	65
		% within measure	6.6%	4.8%	5.7%
	High 2	Count	17	13	30
		% within measure	3.0%	2.3%	2.6%
	Secure 2	Count	17	16	33
		% within measure	3.0%	2.8%	2.9%
	Low 2	Count	5	3	8
		% within measure	0.9%	0.5%	0.7%
Total		Count	573	567	1,140
		% within measure	100.0%	100.0%	100.0%

Reading by year group

Year				Assessment type		Total
				APP	Optional/KS2 test	
3	Reading level	2	Count	26	21	47
			% within measure	26.5%	21.4%	24.0%
		3	Count	65	57	122
			% within measure	66.3%	58.2%	62.2%
		4	Count	7	20	27
			% within measure	7.1%	20.4%	13.8%
	Total		Count	98	98	196
			% within measure	100.0%	100.0%	100.0%
4	Reading level	2	Count	9	11	20
			% within measure	7.8%	9.6%	8.7%
		3	Count	54	34	88
			% within measure	47.0%	29.6%	38.3%
		4	Count	52	70	122
			% within measure	45.2%	60.9%	53.0%
	Total		Count	115	115	230
			% within measure	100.0%	100.0%	100.0%
5	Reading level	2	Count	3	5	8
			% within measure	2.0%	3.4%	2.7%
		3	Count	30	21	51
			% within measure	20.4%	14.3%	17.3%
		4	Count	92	74	166
			% within measure	62.6%	50.3%	56.5%
		5	Count	22	47	69
			% within measure	15.0%	32.0%	23.5%
	Total		Count	147	147	294
			% within measure	100.0%	100.0%	100.0%
6	Reading level	2	Count	1	1	2
			% within measure	0.5%	0.5%	0.5%
		3	Count	17	3	20
			% within measure	8.0%	1.4%	4.7%
		4	Count	121	85	206
			% within measure	56.8%	39.9%	48.4%
		5	Count	74	124	198
			% within measure	34.7%	58.2%	46.5%
	Total		Count	213	213	426
			% within measure	100.0%	100.0%	100.0%

Writing – Whole sample by sub-levels

			Assessment type		Total
			APP	Optional/KS2 test	
Writing sublevel	High 5	Count	5	0	5
		% within measure	0.9%	0.0%	0.5%
	Secure 5	Count	24	31	55
		% within measure	4.2%	7.3%	5.5%
	Low 5	Count	43	0	43
		% within measure	7.5%	0.0%	4.3%
	High 4	Count	64	0	64
		% within measure	11.2%	0.0%	6.4%
	Secure 4	Count	92	173	265
		% within measure	16.1%	40.5%	26.5%
	Low 4	Count	67	0	67
		% within measure	11.7%	0.0%	6.7%
	High 3	Count	83	25	108
		% within measure	14.5%	5.9%	10.8%
	Secure 3	Count	65	112	177
		% within measure	11.3%	26.2%	17.7%
	Low 3	Count	68	32	100
		% within measure	11.9%	7.5%	10.0%
	High 2	Count	35	27	62
		% within measure	6.1%	6.3%	6.2%
	Secure 2	Count	16	17	33
		% within measure	2.8%	4.0%	3.3%
	Low 2	Count	10	10	20
		% within measure	1.7%	2.3%	2.0%
	Below 2	Count	1	0	1
		% within measure	0.2%	0.0%	0.1%
Total		Count	573	427	1,000
		% within measure	100.0%	100.0%	100.0%

Writing by year group

Year				Assessment type		Total
				APP	Optional/KS2 test	
3	Writing level	2	Count	48	45	93
			% within measure	47.1%	44.1%	45.6%
		3	Count	54	51	105
			% within measure	52.9%	50.0%	51.5%
		4	Count	0	6	6
			% within measure	0.0%	5.9%	2.9%
	Total		Count	102	102	204
			% within measure	100.0%	100.0%	100.0%
4	Writing level	2	Count	9	14	23
			% within measure	8.1%	12.6%	10.4%
		3	Count	66	62	128
			% within measure	59.5%	55.9%	57.7%
		4	Count	36	35	71
			% within measure	32.4%	31.5%	32.0%
	Total		Count	111	111	222
			% within measure	100.0%	100.0%	100.0%
5	Writing level	2	Count	5	4	9
			% within measure	3.5%	2.8%	3.2%
		3	Count	47	52	99
			% within measure	33.3%	36.9%	35.1%
		4	Count	73	68	141
			% within measure	51.8%	48.2%	50.0%
		5	Count	16	17	33
			% within measure	11.3%	12.1%	11.7%
	Total		Count	141	141	282
			% within measure	100.0%	100.0%	100.0%
6	Writing level	3	Count	49	56	105
			% within measure	22.4%	25.6%	24.0%
		4	Count	114	132	246
			% within measure	52.1%	60.3%	56.2%
		5	Count	56	31	87
			% within measure	25.6%	14.2%	19.9%
	Total		Count	219	219	438
			% within measure	100.0%	100.0%	100.0%

Mathematics – Whole sample by sub-levels

			Assessment type		Total
			APP	Optional/KS2 test	
Mathematics sublevel	High 5	Count	17	0	17
		% within measure	2.9%	0.0%	1.5%
	Secure 5	Count	21	79	100
		% within measure	3.6%	13.6%	8.6%
	Low 5	Count	44	6	50
		% within measure	7.5%	1.0%	4.3%
	High 4	Count	63	32	95
		% within measure	10.8%	5.5%	8.1%
	Secure 4	Count	77	131	208
		% within measure	13.2%	22.5%	17.8%
	Low 4	Count	95	48	143
		% within measure	16.3%	8.2%	12.3%
	High 3	Count	76	49	125
		% within measure	13.0%	8.4%	10.7%
	Secure 3	Count	89	94	183
		% within measure	15.3%	16.1%	15.7%
	Low 3	Count	58	72	130
		% within measure	9.9%	12.3%	11.1%
	High 2	Count	24	57	81
		% within measure	4.1%	9.8%	6.9%
	Secure 2	Count	15	15	30
		% within measure	2.6%	2.6%	2.6%
	Low 2	Count	4	0	4
		% within measure	0.7%	0.0%	0.3%
Total		Count	583	583	1,166
		% within measure	100.0%	100.0%	100.0%

Mathematics by year group

Year				Assessment type		Total
				APP	Optional/KS2 test	
3	Mathematics level	2	Count	29	43	72
			% within measure	34.5%	51.2%	42.9%
		3	Count	55	37	92
			% within measure	65.5%	44.0%	54.8%
		4	Count	0	4	4
			% within measure	0.0%	4.8%	2.4%
	Total		Count	84	84	168
			% within measure	100.0%	100.0%	100.0%
4	Mathematics level	2	Count	10	29	39
			% within measure	5.7%	16.6%	11.1%
		3	Count	102	96	198
			% within measure	58.3%	54.9%	56.6%
		4	Count	63	50	113
			% within measure	36.0%	28.6%	32.3%
	Total		Count	175	175	350
			% within measure	100.0%	100.0%	100.0%
5	Mathematics level	2	Count	3	4	7
			% within measure	2.3%	3.0%	2.6%
		3	Count	48	51	99
			% within measure	36.1%	38.3%	37.2%
		4	Count	70	59	129
			% within measure	52.6%	44.4%	48.5%
		5	Count	12	19	31
			% within measure	9.0%	14.3%	11.7%
	Total		Count	133	133	266
			% within measure	100.0%	100.0%	100.0%
6	Mathematics level	2	Count	1	0	1
			% within measure	0.5%	0.0%	0.3%
		3	Count	18	27	45
			% within measure	9.4%	14.1%	11.8%
		4	Count	102	98	200
			% within measure	53.4%	51.3%	52.4%
		5	Count	70	66	136
			% within measure	36.6%	34.6%	35.6%
	Total		Count	191	191	382
			% within measure	100.0%	100.0%	100.0%

Annex

This independent evaluation report was produced during the pilot of the assessing pupils' progress (APP) approach between 2006 and 2008. Interim evaluation reports were also produced during the pilot, which informed the work of the APP development team, enabling the team to make adjustments and refinements in response to feedback from pilot schools.

This report refers to some features that were specific to the pilot phase and will not apply to the APP approach when used by schools beyond the pilot. In particular:

- **Data from a sample of pupils in the class:** To systematically collect and analyse data from pupils in the pilot schools, the key stage 2 pilot required participating teachers to submit termly data from a sample of between 6 and 12 pupils in their classes. Most schools continued to provide assessment focus level data on these pupils during the second year of the pilot. Some teachers started to use the assessment guidelines with all of their pupils before the end of the pilot, but this was not a requirement of participation. We anticipate that, when implemented as a mature system across a whole school, teachers would apply the APP approach to all pupils. There are materials to support this process on The National Strategies website at www.nationalstrategies.standards.dcsf.gov.uk/primary/assessment/assessingpupilsprogressapp.

- **Moderation visits:** It was critical to the evaluation of the pilot to obtain a clear picture of the types and quality of evidence that teachers used to inform their judgements. It was also important to have direct support available to the pilot schools and teachers so that they could clarify issues and get help using the approach and materials consistently. This meant that, in the key stage 2 pilot, a number of experienced, independent associates were deployed by QCA to visit the pilot schools each term to look at examples of teachers' judgements and some of the evidence they took into account. This method of moderation was particular to the pilot. We are keen to find out about standardisation and moderation practice as it develops at local level. QCA, DCSF and the Chartered Institute of Educational Assessors will continue to look at models and options for supporting this activity.