

Strengthening Teachers' Summative Assessments:

**Report to the Nuffield Foundation
from King's College London & Oxfordshire LEA**

Report prepared by:

Dr Chris Harrison
Dr Jeremy Hodgen
Professor Paul Black
Dr Bethan Marshall
Ms Natasha Serret
Ms Dorothy Kavanagh

Report submitted:

December 2007

STRENGTHENING TEACHERS' SUMMATIVE ASSESSMENTS

The King's-Oxfordshire-Summative-Assessment-Project (KOSAP) was an 18 month project, from March 2006–September 2007, which aimed to investigate how to help teachers enhance the validity and reliability of their assessments so that these can play a significant and trustworthy part in all summative assessments of their students. This was a collaborative development between teachers in Oxfordshire schools, their schools' managements, LA advisers, and experts in school assessment at King's College London (APPENDIX F). It involved investigation of the possibilities and practicability of assessment of year 8 (Y8) pupils within the domains of English and of mathematics in three Oxfordshire secondary schools. A pilot study, funded by the DfES, ran from January 2005-March 2006 (APPENDIX E) and focused on an exploration of existing summative practices of two English and mathematics teachers from each of the three schools and development and trialling of new or adapted assessment activities. This current phase of the research focused on how the teachers implemented these assessments within their own departments. Key research foci were the constraints and affordances that arise as teachers take a more active part in designing, using and evaluating summative assessment tools and activities and the subsequent effects in the classroom as the formative-summative interface is brought closer together.

OUR APPROACH

The intention of this research was to work with teachers to survey their initial assessment tools and practices, to appraise how they judged and valued these, and to see whether and how they could be supported in improving them. We wanted to do more than simply evaluate whether teachers could implement assessments that were provided for them. Rather, we wanted to understand the ways in which they interlaced assessment with curriculum and pedagogy through allowing them to explore for themselves how they might develop and evolve better teacher assessment.

In the light of these intentions, our initial plan was to learn in detail about the current assessment practices in the three English and three mathematics departments, to explore the rationale for these, and then in the light of the findings, to suggest changes which would help them develop and improve. At the same time we hoped to develop with the teachers a dialogue, both within and across their schools, as a basis for inter-school moderation work. A second stage of this plan was to help schools, through critical reflection, to take on some changes and to produce evidence of the quality of their improved summative assessments. Through working in English and mathematics, we thought we might develop different perspectives on this task which might give us insights for future development within these subjects and perhaps for other curricular areas. Out of this we could begin to propose procedures which would both be feasible and effective in ensuring high quality assessments.

This project provided the impetus for teachers to investigate three issues, namely the role that assessment takes in terms of how they make judgments of students' attainment, why and how these decisions foster or limit learning experiences in their classrooms, and how they deal with the pressures they encounter when sharing assessment information with parents, students, teacher colleagues, senior leadership teams, and OfSted. It has enabled the project teachers to consider what the 'ideal' assessment situation might be for them and to see how far they might reconcile the achievement of this ideal with the policy constraints and practical realities that they face on a day-to-day basis in their schools. The problem lies in unravelling the complexities of this reconciliation to produce a workable system of teacher assessment that has sufficient rigour and validity.

For an overview of project meetings, see APPENDIX A, and for an outline of our methodological approach, see APPENDIX B.

THE TEACHER ASSESSMENT SYSTEMS

Both the mathematics and English teachers produced portfolios of work for each of their Y8 students over the course of the year. The teachers drew on their existing experiences and skills in both summative and formative

assessment, in particular coursework moderation at GCSE, whilst also questioning these current practices. In mathematics, developing suitable assessment tasks proved to be significant piece of work and, as we describe in Appendix D, several tasks were piloted then later dropped as unsuitable. This gradual development of the tasks, and the discipline of formulating agreement on some common tasks between the schools, forced their assessment tasks to be completed by their students in a few months, while the English teachers spread their tasks throughout the year. This led to some of the mathematics tasks becoming ‘bolted on’ rather than embedded within the curriculum and the teachers recognised that this was not ideal.

In deciding the balance and number of final assessments, there were two competing concerns: the need for the assessment to be manageable for teachers and the need to provide sufficient evidence.

For English, it was decided that work from writing, reading and speaking and listening should be assessed, three for each strand. At least two of these assignments would overlap and one of them was completed in controlled conditions. The actual tasks and contexts varied between schools. (See APPENDIX C for a detailed description and discussion of the English tasks)

In mathematics, students were assessed by six tasks. Three tasks were common to all the schools, whilst each school assessed students using an additional set of three tasks. The tasks assessed both process and content skills and covered all the attainment targets. (see APPENDIX D for a detailed description and discussion of the mathematics tasks.)

In both subjects, it is clear that assessment impacts on both pedagogy and curriculum. For this group of teachers, re-thinking assessment highlighted both strengths and weaknesses in their understandings of and practices in pedagogy and curriculum. Summative assessment was successful where teachers had experience: the English teachers were better at assessing writing and the mathematics teachers at GCSE-type coursework activities. In both cases, they provided richer and more detailed assessments, their judgements covered a narrower range and they reached consensus in moderation meetings more quickly. The work was much less successful in areas or topics where the teachers had limited previous experience: both reading *and* speaking and listening for the English teachers, Shape, space and measures for the mathematics teachers, and classroom talk for both groups.

I remember, there was one quite heated debate that we had about kind of task, about how you assess reading and whether you could assess reading through (pause) assessing reading through writing, assessing reading through speaking and listening. So whether you are assessing those skills separately or whether you can assess them together.
(English teacher, November 2006)

However, some of the mathematics teachers felt that designing assessment tasks for Shape, Space and Measures, an area for which they had less expertise, enabled them to be more radical and original (June 2007 interviews). In particular, the unfamiliarity of the task encouraged them to consider the assessment criteria in more depth and to avoid simply adapting GCSE coursework criteria.

In both subjects, teachers encountered student motivation issues. Students did not appear to take “new” assessment tasks as seriously as the more familiar and recognised assessment activities (traditional tests in mathematics, essay writing in English.) The introduction of a new form of assessment requires student support, particularly when “students like tests” (English teacher, November 2007). From the small sample of student group interviews, it appears they were not aware that anything radical was going on in terms of assessment. However, they did talk favourably about some of the KOSAP assessment tasks that they had experienced but these were seen as “good lessons, enjoyable [and] rewarding activities” (Student interviews, June 2007). But, if teacher assessment is to be successful, students need to be involved and aware of the changes in order that they produce their “best” work.

Part of the work in the latter part of the project involved disseminating the ideas to other colleagues in their departments and involving these colleagues in implementing the assessment tasks that had been developed. For moderation purposes, a sample of these portfolios were then ‘blind marked’ in each of the English and

mathematics departments of the three schools and brought to the cross-school moderation meeting in June 2007. Schools 'blind marked' samples from the other to schools and aggregated levels were agreed through discussion for each portfolio. Moderation across schools was generally successful and the project teachers were, in the majority of cases, able to agree on specific National Curriculum levels for each of the portfolios that they examined.

Both subjects had difficulties with the assessment criteria. For the English teachers, the National Curriculum levels are too vague, whilst GCSE criteria are too sophisticated for the quality of writing that most 13 year olds will produce. Added to this, the disjuncture between the KS3 criteria and those for GCSE was an added difficulty that the teachers needed to grapple with. For both subjects, there was a tension between those who want very specific criteria and those who favour a "looser" set of criteria allowing more room for interpretation by teachers and students; some mathematics and some English teachers wanted a tighter set of criteria, whilst others were concerned that specificity would result in losing the important, bigger, holistic picture. Neither subject has resolved these difficulties.

DISCUSSION OF FINDINGS

In this section, we discuss the main issues that we encountered. This is presented as a set of five classes of obstacles and problems: whilst this approach may at first appear as pessimistic and defensive, the account will show how the problems were tackled and how teachers' came to understand them and to achieve success in tackling them. This thereby provides a basis for the positive recommendations made in our closing section.

1 External tests and curriculum

It was clear that we underestimated the pervasiveness of tests. This is particularly strong in mathematics, but there is also some evidence of this influence in English. In mathematics, even where teachers were intent on introducing an alternative system of teacher assessment, they retained the use of formal tests, particularly KS3 tests, as examination practice and preparation. But, if such tests are used, the results inevitably compete with the teacher assessments. Most teachers felt this pressure at least as a strong constraint, or even, for some, as an over-riding determinant:

The real assessment, the proper assessment is the one they do in the exam. I think that this is a fairly common view, because it's the one that goes on the certificate and it's the one that makes a difference. It's the one that gets them into college or gets them a job interview. It doesn't matter if they have done brilliantly for five years of school. They do badly on one day and they get a D on a piece of paper and they are stuck.

(Mathematics teacher, May 2005)

The coursework requirements of the statutory testing system were also unhelpful and there were also problems arising from the weakness in the National Curriculum specifications. As one teacher expressed this:

If a portfolio system was to be put in place, the current level descriptors are woefully inadequate and do not lend themselves to consistency. (English teacher, November 2007)

In both subjects, teachers' own assessments were marginalised at Key Stage 3 by the focus on the external tests. A further problem was created by the inconsistent demands of, on the one hand, Key Stage 3 testing and, on the other, the GCSE criteria – a serious problem in English particularly in relation to assessment of speaking and listening. In GCSE mathematics, the external requirements had created stereotyped practices in using and applying mathematics which undermined the purpose and diminished the effectiveness of the teacher assessment and this affected how the teachers viewed their assessment tasks for Key Stage 3. The quotation above illustrates how project teachers had to think more carefully than hitherto about the curriculum statements against which these were to be categorised: they often found, in both subjects, that the statements were inadequate whether for this purpose or for informing their task of steering the progression of pupils.

2 School policies and parental expectations

Although we had ensured schools' commitment to the project, by initial visits to discuss the aims with each of school's senior leadership teams, by contact with key members of these teams, and through the local authority personnel, it became clear that the developmental work that the English and mathematics teachers were not used to direct the school's assessment agenda. This caused problems for the project teachers because the expectation was that they were to continue with their current system whilst, at the same time, attempting to change it. Such problems arose because, although the schools' leaderships accepted at the outset our explanation of our planned intentions, they did not grasp the implication that KOSAP would result in different assessments of students. So, they understood the need to enable teachers to do KOSAP work in school time, but not that this would result in different kinds of information being produced. Our own initial presentations to them of the implications of the work were also inadequate because we were not ourselves aware of the size of the gap between the schools' practices and the qualities of summative assessment we were hoping to develop. For example, the heads and deputies understood the need to enable teachers to do KOSAP work in school time, but did not anticipate that this would result in different kinds of information being produced, for reports to senior management, reports to parents, and for year-to-year pupil records.

Teachers felt the pressures of the external tests system, in part through the priority given to the published test results by school managements and by parents and pupils, both to achieve in these tests, but also and to report on learning in terms only of a single level or grade.

You go to a parent's evening. I will tell them what level they are working at and they ask, 'but what will they get in the test?' As if somehow this was far more reliable than my teacher assessment.

(English teacher, May 2005)

We are caught in a trap. Report to parents a level. Students focus on just doing tests so we have a level to report. We have missed the point of what we are trying to achieve. It is less and less with monitoring progress and more about filling in boxes.

(Mathematics teacher, May 2005)

In one of the three schools, these constraints caused particular difficulties for the KOSAP teachers because the school introduced a 2 year KS3 (over Y7 and Y8 in place of Y7, Y8 and Y9) and, in consequence, had pupils tackling the KS3 examinations in Y8 alongside the teacher assessment in KOSAP. Inevitably, the KS3 examinations dominated.

Throughout the project, the teachers were aware of the issues we were exploring and in the improvements we were suggesting and, while attempting to bring about change, they felt constrained by their departmental, school and wider contexts. These teachers were committed but had (1) limited time *and* insufficient opportunity to develop their expertise and (2) limited space within school and subject cultures to assess pupils in different ways. There is a very real difficulty in that success within school cultures is framed in terms of success in examinations rather than success in learning. The status quo option is very powerful both because effecting any change is difficult and because the existing assessment system has such high credibility at all levels in the assessment cycle and with different audiences.

3 Competing claims and agendas

Some of the teachers themselves had agendas which made it difficult to engage fully with the project's work. In mathematics, for example, one school was concerned with two objectives: introducing a system of assessment that more closely aligned with KS3 and GCSE, and ameliorating the effects of a group of weak teachers. In English, two of the schools strove to empower learners by involving them much more within the summative assessment process and, while this was a laudable aim, concentration on it distracted the English teachers away from the wider agenda of the project.

Other teachers had their own personal agendas – notably new jobs and starting a family¹. The project's aims required far more change from the teachers than we envisaged, and so they were more difficult to achieve, within the normal working of schools, than we had anticipated. The issue was not time *per se*, although certainly more time would have helped. Of relevance here is Ruthven's (2007) notion of the time economy in schools; for these teachers, in a sense, there were always more pressing priorities in school than the KOSAP work, so creating regular space to think and talk was difficult, even where time was available: this militated against the profound changes required in re-thinking assessment principles and relating these to reforms in pedagogy and assessment practices.

4. Teachers' assessment literacy and confidence

In our initial exploration of the existing assessment practices within English and mathematics in the three schools we soon realised that the application of the principles that guide quality in assessment, notably the concept of validity, seemed weak – more so amongst the mathematics teachers than amongst the English teachers. It was not, however, difficult to involve them in a debate, notably starting from the question “What does it mean to be good at (English or maths?)”. The teachers readily engaged in this debate, and through such engagement began to see that they had, in their practice, neglected to critique their own work in the light of their core beliefs and values concerning the purpose of learning in their subject.

Overall, the perceptive application of concepts of validity and reliability, to the extent of taking these seriously in auditing their own work, was seriously in need of development amongst the teachers. Teachers came to realise the importance of these concepts – their problems in relating them to practice are illustrated in the following reflections expressed by two of them:

It is important that teachers understand the skills that are being assessed (and make up what it is to be good at English) rather than being fixated on the form of the assessment. Teachers can then communicate the value of the assessment rather than just focussing on the test in isolation.

(English teacher, November 2007)

It all points towards the ‘what does it mean to be good at maths’ question and how we give (get?) the students to show this – surely tests in a formal way (if properly constructed) have a role to play in allowing students to demonstrate this – and does also leave scope for teacher assessment – if the teachers are confident in this.

(Mathematics teacher, November 2007)

We have made significant progress in stimulating work amongst these teachers to both recognise and start to meet all of these needs within their practice. Two teachers described this progress as follows:

I think, that, I think I was almost falling into a very checklisty trap, The only thing I could say is the assessed criteria and then you jump through these hoops and I knew it was a flawed thing. But I couldn't articulate why and now I've got a language to talk about that and it kind of holds a new set of discourse about assessment that I didn't have before and I found that really, really valuable in kind of understanding you know some of the chaos.

(English teacher, June 2007)

I knew I hated the KS3 tests and found the preparation of students for these tests as one of the most unrewarding aspects of my job. However, through the project, I now have a much better understanding of why the tests are so problematic.

(English teacher, November 2007)

The fact that this needed to be done is confirmation of John Gardner's (2007) view of teaching as ‘a partial profession’ in that it is weak in one of the main requirements for professional teachers – assessment literacy.

¹ In total, we worked with 10 mathematics teachers and 8 English teachers.

Linked to this was the generally conservative attitude that the teachers have towards the task of making summative judgments, and this was recognised by the project teachers. There is a general acceptance of the tests and tasks that they already do, despite their concerns that these assessment tools may be unfair, invalid, and unreliable in measuring the capabilities of their students. They recognise that the tasks and tests they set are poor and do not help them to identify the key problems of their pupils. This hinders both the improvement of pedagogy and the development of better summative assessment.

Things I became more interested in was the poster, oral presentation, (students) doing powerpoints. I wonder if we forced ourselves back into coursework mode. Maybe these (assessment tasks) take too much time to work out or too much creativity.
(Mathematics teacher, November 2007)

5 Problems in the current practice and culture

The next problem, identified in our initial exploration, was that even within its own framework the teacher's existing practices lacked the rigour and uniformity required. Thus, before the teachers were able to begin working on improving quality of existing assessments, we spent a lot of time in the project working with the teachers to identify and understand these problems and build their confidence, both in the procedures and in the tools available to them within the current systems.

There is also a lack of confidence on the part of teachers in the school (in maths) in their own judgment. Levels from tests are trusted whereas teachers' assessments are not and often a test mark agreeing with a teachers assessment is seen as vindication rather than the other way round. . . . Student perception of teacher assessment seems to be that it is a 'guess' which is then confirmed or otherwise by the test – this culture would take a lot of changing.

(Mathematics teacher, November 2007)

Thus, for some of the teachers, there was a concern from the start about whether it was possible to produce school-based assessments that would be seen as acceptable to the parties involved. There was an uncritical acceptance of the systemic external constraints in testing, even when they would agree that these did not represent their own values.

One aspect of the project's approach in addressing these in practice was to help teachers develop their own critique of 'off-the-shelf' test questions and so to produce for themselves questions of better quality, more attuned to their own teaching schemes and values:

The maths department of (School C) has in place a series of end of unit tests which, at the time, generated a series of numbers without any relation to areas of the national curriculum. The first thing that Bill and I did was to rewrite the tests in order that they generate useful information about how well students are doing in each area against NC levels.
(Mathematics teacher, November 2007)

The project made me think more critically about what exactly I was assessing. The first question I remember being asked ('what does it mean to be good at English?') gave me a different perspective on assessment I find myself continually returning to this question. When I began teaching, I accepted that tests (were) established by the department with reliable methods – e.g. the timed essay for assessing a student's reading of a text. The question has led me to question this assessment of reading in particular, since I have realised that the essay form can be more about the students' writing ability, than their reading.
(English teacher, November 2007)

The task of developing the critical appreciation of assessment instruments is a demanding one, for it makes strong demands on any teacher's subject knowledge and pedagogical content knowledge. This appeared to be a greater problem for the mathematics than for the English teachers.

Alongside this development of discernment and skill in relation to formal testing was a development in the practice of crafting, applying and assessing classroom tasks so that each school could build up a portfolio on the basis of which the summative assessment of each pupil could be determined. This had to be achieved through

collegial endeavour in constructing tasks which would be valid and which would allow opportunity for pupils at all stages of development to do themselves justice. These skills, first developed in the project's inter-school meetings, were then used for corresponding developments within each school department:

Implementation of tasks – staff were a bit reluctant to do the project... but post-project their views changed and this year development of investigation based tasks has become an issue that the KS3 staff have been keen to do ...
(Mathematics teacher, November 2007)

An issue about development of open-ended tasks that was particular to mathematics was the 'local / global' issue: the mathematics teachers had difficulty seeing the mathematical commonalities in different tasks, so they were rather rigid and inflexible in allowing interpretation of tasks (either by other teachers or by pupils).

However, several teachers saw that the changes had led to a new understanding by pupils of their participation of their work for assessment:

...helped me to think (and enable others to think) much more deeply and critically about how to set quality tasks. Also enabled me to share success criteria in a way which is more valuable and engaging and apprentices students into the 'guild knowledge' that subject specialists have. Project has removed anxiety about delineating success only in terms of a neat, prescriptive check list.
(English teacher, November 2007)

How the investigation tasks have improved after the first one or two and the pupils would develop their thinking with little or no teacher intervention choosing their own approach to a problem. This highlights the need for them to be appropriately placed in the scheme of work for KS3 to link in with topics covered.
(Mathematics teacher, November 2007)

This development of quality with individual tasks was a first step in meeting the essential requirements for ensuring quality in the practice of moderation within schools, namely the development of assessment literacy amongst all staff and the achievement of fairness and uniformity in summative assessments of pupils.

Inevitably, as the portfolio components were improved, there had to follow a development in the procedures for moderation. The experience of the English teachers with the teacher assessment component of GCSE was helpful here, but even for them new lessons were learnt:

... the moderation and standardisation process was incredibly valuable in ensuring rigour, consistency and confidence with our approach to assessment. ... teachers in the school were highly motivated by being involved in the process that would impact on the achievement of students in their classes (English teacher, November 2007)

There was also an issue of the lack of confidence, on the part of many teachers, in their professional judgments - which they realised depend essentially on their expertise as teachers. This was more evident generally across the mathematics teachers, although apparent, to some extent, for the English teachers. Lack of communication about certain aspects, and opportunity to share practice, as well as confidence in completing the assessments, was also more evident in the teachers who had not had the experience of project meetings,

So to be honest I have no idea.... how other people assess speaking and listening, So that's maybe something I need to get in around this. So you know, maybe observe some speaking and listening.
(English teacher, November 2007)

and this was attributed to the lack of developmental opportunities in contrast to the those that the project meetings had provided for the KOSAP teachers.

I wish they (departmental colleagues) could start from the beginning just like we did. It's hard to take on this way of doing things if you haven't been through what we did.
(English teacher, June 2007)

In general, in this as in other respects, the task of disseminating across departments the assessment expertise developed in the project was not an easy one. As two teachers explained:

Some of our staff just don't get it. They're OK with assessment for learning . . . They know we need to do tests. But they don't really link the two. (Mathematics teacher, November 2007)

I wish we could bottle what we've done. Our teachers are happy to go along with this, but it's so hard to help them deal confidently with the issues that arise for them in their own classrooms without . . . You know . . . the hindsight that we have through engaging with this. (English teacher, June 2007)

With respect to the project's more ambitious aim, namely to work towards synergy between the formative and summative purposes of teachers' assessment work, it was soon obvious that progress could not be made until the quality of the practices which served both of these purposes could be improved. Teachers had to formulate agreed ways to resolve the tension between giving formative help to pupils to improve their work on an open-ended task and at the same establishing and implementing criteria of fairness for using the final products to assess these pupils. The design and trialling of tasks conducted under controlled conditions was a significant step here, for teachers could give pupils strong formative guidance in their preparation for such a task, leaving pupils entirely on their own to show how they had benefited from such learning opportunities, and yet maintaining close and productive synergy between the learning and the assessment.

What we did not find, in the pursuit of this work, was any strong concern amongst the teachers about possible corruption of the assessments through plagiarism or through undue help so that work assessed was not the pupils' own. Teachers were confident that they knew their pupils sufficiently well that they could detect, or easily check, that work was not their own. Of course there could be help given, say at home, but if this helped the pupils to develop their own skill and understanding, it would have to stand as an objective achievement.

However, in relation to teachers' regard for the work of some of their colleagues, we did see some signs of mistrust, which appeared at a number of levels. One of the opinions coming from some of the teachers (across both subjects) was that they could only fully trust those with whom they had direct contact. There was also a need for trust amongst teachers so that they could accept, in meetings, to disagree or to be "wrong" – even the "best" measurement instruments are unreliable to some degree and hence teachers are bound to make assessment errors. Overall, this points to the need for collegial development of instruments and of moderation procedures so that trust is built through co-operation, and so that that unfair practices by teachers would be unlikely to arise, or to go un-detected.

RECOMMENDATIONS

Introducing summative assessment by teachers is a formidable but necessary objective

Our experience shows that the improvement of teachers' own in-school assessment skills so that they are fit for the purpose of guiding the progress of pupils is a formidable task. We stress that there were many positive benefits from this project, in particular all the teachers felt that they had made considerable changes to their practice. That this was such a positive and keen group of teachers highlights both, the difficulty and urgency of the problem. A second and more ambitious aim, of establishing the quality of teachers' own summative assessments so that they may claim to supplement or even replace formal tests externally set and marked (ARG 2006), will be even more difficult to achieve. Nevertheless, our study adds further weight to the imperative for teacher assessment.

Continuing Professional Development (CPD)

As we have discussed above, our research has demonstrated considerable weaknesses in the assessment literacy of teachers. It is clear that there is an urgent requirement for CPD to strengthen this. As the teachers we worked with found, this will have considerable additional benefits beyond assessment in improving teachers' pedagogy and understanding of the curriculum.

Establish strong collegiality in assessment

We believe that there are many advantages to organising this CPD within schools in order to develop strong collegiality within and between subject departments. However, whilst there are considerable advantages to teachers themselves collaborating, teachers do need support from specialists within schools and outside experts. Re-thinking and developing assessment practice takes time and teachers will need time to do this as individuals, groups and departments. However, we do not think that the answer simply lies in providing teachers with additional time. Making fundamental long and medium term changes to teaching are difficult to balance with many of the short-term demands of teaching. The teachers that we worked with strongly valued the pressure provided within the project and school managements need to similarly give a high profile to work on assessment. An agenda for such CPD would include:

- Establish agreed departmental criteria for validity.
- Agree the optimum balance for them between formal tests, ‘controlled conditions’ tasks and tasks conducted through the normal classwork timetable.
- Review test instruments currently used and work to improve ‘off the shelf’ tests and tasks
- Agree guidelines for the formative/summative interface in relation to formative help given on tasks to be assessed.
- Develop understanding and transparency with pupils about all assessment procedures

Build on and extend teachers’ existing assessment strengths

It is clear from this research that teacher assessment was strongest in areas where teachers already had experience and expertise, but transferring this expertise was not straightforward. Hence, teachers need help and support to do this. We recommend that teachers should spend time describing, developing and critiquing both criteria and exemplars of student work. Developing multiple but “equivalent” sources of evidence facilitates agreement between teachers. A further benefit is that such rich description enables teachers to describe and, thus, transfer their expertise.

Work with local clusters of schools to exchange tasks, alignment standards, and thereby strengthen CPD

Schools do not and cannot work in isolation. Working with teachers and departments from other schools provides an obvious opportunity to compare, contrast and “test” approaches. Equally important is the building up of trust beyond the confines of a single school. Trust provides a context in which teachers can disagree in order that they can develop what Wiliam (1998) refers to as *construct referencing*, or a shared understanding of the what, why and how of assessment. A particular focus of this collaboration between schools should be to negotiate and foster inter-school cluster arrangements for local moderation as an aid to CPD and gaining public confidence in and for all schools.

The role of school leadership

Overall, what the project was encountering was a need to challenge the uncritical acceptance of current practices from school leaderships. There had to be a challenge about the quality of existing instruments and practices. This not only weakened the learning work with pupils, but could also weaken the decisions and guidance about their progress in school, notably because of the practice of reporting about individual pupils, in their transition from one school year to the next, in a single aggregated grade or mark, and thereby ignoring the rich data teachers can have of the uneven profile of strengths and weaknesses in the performance of any one pupil. It is the responsibility of management to do something about this.

Appoint and train an assessment leader in every school

This would ensure that at least one teacher within each school had some assessment literacy. This teacher would have two main roles: first, to act as an expert to help curricular areas re-think assessment, and second, to provide a critical perspective on assessment to inform school leadership teams. For future developments in assessment, it is essential to involve schools’ assessment leaders in the development and to involve them both in helping teachers balance teacher assessment with statutory testing and in finding ways of creating space within the school assessment system for teacher assessment to develop.

The system of assessment at KS3 and GCSE

We are conscious that a revised National Curriculum has been just introduced (QCA, 2007) and that significant changes are currently either underway or being developed at GCSE (e.g., Functional Skills assessment, a second GCSE in mathematics.) It is important to allow time for such changes to “bed down”. Hence, we make just two recommendations concerning national systems of assessment that our research has shown to be particularly urgent.

KS3 national tests and GCSE English should be better aligned

In English, the mis-alignment of the national testing regime at KS3 and GCSE is a serious problem. There is a need to develop KS3 English as a stepping stone towards progress towards GCSE and to evolve a set of criteria that evolves smoothly from one key stage to the next.

GCSE mathematics assessment should include some extended tasks assessed by teachers

The antipathy of mathematics teachers to the current form of GCSE coursework is well-documented (e.g., QCA, 2006) and this view was largely shared by the mathematics teachers in our study. We are encouraged that the revised National Curriculum places considerable emphasis on mathematical processes. However, without some element of coursework at GCSE and at KS3, we strongly believe that these good intentions will not be realised. Our work suggests that referring to these as “extended tasks” rather than “coursework” would be beneficial, and that “controlled conditions” would allay many of the reservations mathematics teachers have.

FURTHER RESEARCH AND DISSEMINATION

At several points the project’s work has revealed problematic aspects of the interacting effects of assessment with both curriculum and pedagogy, often controlling in deleterious ways the link between curriculum aims and the pedagogy that should give priority to their implementation. It seems clear that a key feature of the work of teachers is the way in which they handle the mutual interactions of curriculum statements, assessment pressures and practices, and their pedagogy. In the best outcome, teachers achieve a benign equilibrium of reconciliation between these. More often they struggle to balance the contradictions between them. Changes in any one component, e.g. between a curriculum specified loosely and one specified in great detail, or a change in the external testing, or a drive to implement formative assessment, can challenge the existing equilibrium, and either enhance or alleviate the contradictions. The history of national and local initiatives to improve teaching and learning is replete with examples of unintended consequences, which arise because the complexity of these interactions is not understood. The studies of the ASLO and ARIA² project of the Assessment Reform Group will be an important contribution to our understanding of these problems, but we believe that more in-depth empirical study, that can build on their findings, will be needed if better guidelines can be formulated for policy change in the future.

We have already begun disseminating this research through an AERA, a BSRLM and two BERA conferences; we were also invited to give presentations to the ASLO ARIA and SQA groups.³ We have plans for publications for research and professional journals and also for writing a booklet for teachers. We hope to organise a seminar in 2008 to bring together key personnel from government, examination boards and other expert assessment groups to share and discuss our findings. We wish to discuss the next stage of dissemination with foundation officers in the near future.

² ASLO: Assessment of Significant Learning Outcomes; ARIA: Analysis and Review of Innovations in Assessment.

³ AERA: American Educational Research Association; BSRLM: British Society for Research into Learning Mathematics; BERA: British Educational Research Association; SQA: Scottish Qualifications Authority

REFERENCES

- ARG (2006) The role of teachers in the assessment of learning (London: Assessment Reform Group, Institute of Education).
- Black, P., & Wiliam, D. (2006). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning* (pp. 119-131). London: Sage.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting It into practice*. Buckingham: Open University Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2006) *Strengthening Teacher Assessment Practices, Learning and Evidence*. Paper given at BERA University of Warwick
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2007) *Strengthening Teacher Assessment Practices*. Paper given at AERA , Chicago
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2007) *Riding the Interface*: Paper given at BERA Institute of Education, University of London
- Brown, M. (Ed.). (1992). *Graded Assessment in Mathematics (GAIM)*. Walton on Thames: Nelson.
- Crookes, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education*, 3(3), 265-285.
- Gardner, J. (2007). Is teaching a "partial" profession? *Make the Grade*, Summer 2007, 18-21.
- Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal*, 16(2), 153-176.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26
- Kelly, A. E. (2003). Theme Issue: The Role of Design in Educational Research. *Educational Researcher*, 32(1), 3-37.
- QCA, (2006). *GCSE Mathematics Coursework: Consultation Summary*. QCA/06/2737. London: Qualifications and Curriculum Authority.
- QCA, (2007). *National Curriculum*. Retrieved 11/12/2007, from <http://curriculum.qca.org.uk>.
- Ruthven, K. (2007). *Teachers, technologies and the structures of schooling*. Plenary lecture at CERME 5: Fifth Congress of the European Society for Research in Mathematics Education, Larnaca, Cyprus
- Wiliam, D. (1998). *Construct referenced assessment of authentic tasks: alternatives to norms and criteria*. Paper presented at the 24th Annual International Association for Educational Assessment (IAEA) Conference, Barbados, West Indies.

APPENDIX A: Overview of project teacher meetings

Project meeting date	Details
1st March 2005 (half-day)	Introduction to the project, KOSAP teachers and Kings and Oxford team
11th May 2005	Sharing existing assessment practices Discussion about the nature of assessment tasks, teacher judgement, levelling and teacher intervention within the assessment process. Reflection log
29th June 2005 (English moderation day)	Teachers and KOSAP team attempt an English writing task Departments bring sample of student portfolio of work for yr 8 for moderation
6th July 2005 (Maths moderation day)	Departments bring sample of student portfolio of work for yr 8 for moderation
13th October 2005 (half-day, whole KOSAP group)	What does it mean to be good at English/Maths? 2.What counts as evidence to justify this judgement? 3.Does the nature of the evidence change depending on audience? 4.What are doing as part of this project in terms of assessment in the light of these questions?
22nd February 2006	What is the prime reason for each assessment? Who is the audience for the assessment information? What does the assessment data need to be like to fulfil this purpose? How confident do we need to be in the assessment data? Should the assessment look backward or forward? Followed by subject group moderations

	<p>(including videos for EN1 in English)</p> <p>Introduction to aspects of Crooks et al use of chains to highlight key components in the assessment process</p>
25th May 2006 (English moderation day)	<p>Prior moderations in departments. These and a final agreed level for each candidate was brought to the day for further moderation within the KOSAP group. Issues arising:</p> <p>Weighting between EN1,EN2,EN3 KS3 or GCSE criteria?</p> <p>Methods and rationale for aggregation</p> <p>Judging less tangible aspects (flair)</p> <p>Reliability and validity of tasks, use of controlled conditions</p> <p>Teacher reflection log</p>
14th June 2006 (Maths moderation day)	<p>Prior moderations in departments. These and a final agreed level for each candidate was brought to the day for further moderation within the KOSAP group. Issues arising:</p> <p>Interpreting criteria</p> <p>Aggregation</p> <p>Reliability and validity of tasks</p> <p>Teacher intervention</p> <p>Manageability (time)</p> <p>Teacher reflection log</p>
27th June 2006 (Action planning day)	<p>Development of English portfolio for next phase in KOSAP</p> <p>Concerns in maths about the quality of maths teaching and the impact this has for using class work for assessment. The influence of testing on this.</p>
27th November 2006	<p>Presenting and standardising the tasks for assessment</p>

	Approaches needed for wider dissemination across the department- truncating the KOSAP experience for other teachers
19th September 2006 (Maths twilight)	Further work on developing and standardising maths assessment tasks
5th March 2007	Report on the status of the assessment tasks Negotiating moderation procedures (sample size, aggregation, blind marking)
May 2007 (English moderation day)	Portfolios for different departments sent to schools beforehand. Blind marks brought to the day. Discussion around a selection of the portfolios. Teacher reflection log- reflection on the tasks
July 2007 (Maths moderation day)	Portfolios for different departments sent to schools beforehand. Blind marks brought to the day. Discussion around a selection of the portfolios. Teacher reflection log-reflection on the tasks
November 8th 2007 (Whole group final project day)	Teacher Reflective Writing Discussion of teacher learning from project in separate and then mixed subject groups. Teachers devise recommendations.

APPENDIX B Data Collection: Methodology

The project has used a design research methodology (Kelly, A, 2003) within a mixed methods approach (Johnson & Onwuegbuzie, 2004) adopting and extending methods of data collection and analysis used in our previous research.

Data sources include:

Regular interviews with the 12 core teachers- these 12 core 'KOSAP' teachers were interviewed three times over the duration of the project (May 2005, November 2006, July 2007) - and a small sample of other, 'non-KOSAP' maths and English teachers working within the departments were interviewed in November 2006 and July 2007 in order to gauge the extent to which ideas and approaches developed in the project had been disseminated more widely throughout a department;

Pupil questionnaire- the same questionnaire was carried out twice on the same groups of year 8 students, in October 2006 and again in July 2007. The questionnaire was distributed to all year 8 students in the three participating schools and explored student attitudes to learning and assessment in the context of English and maths;

Focus group interviews with small groups of year 8 students, selected by the core teachers, were carried out in order to enrich the data generated from the questionnaire;

Fieldnotes of all whole group and subject group project meetings;

Teachers' writing and reflections, carried out during group project meetings;

Evidence of Year 8 summative assessments in English and maths submitted for project moderation days;

Fieldnotes and audio recordings of school departmental moderation meetings, as well as inter-school meetings preparing for these in relation to Year 8 English and maths assessments conducted as part of the project;

Classroom observations of summative assessment events.

APPENDIX C: ENGLISH ASSESSMENT TASKS

In English, the assessments assessed reading, speaking and listening, and writing. The actual contexts for tasks varied between the three schools, although all the tasks went through an extensive standardisation process. All three schools set a task under controlled or timed conditions and all set a task intended to assess both reading and writing. For all tasks, the final piece of assessment was the culmination of an extended unit of teaching. Specific descriptions of the tasks /contexts in each school follow a discussion of the general issues relating to the three National Curriculum strands.

Speaking and Listening (En1)

Tasks: Structured speaking and listening activities

Assessment: Using modified GCSE Speaking and Listening assessment criteria. The groupwork involved some peer assessment using these criteria.

Commentary: This task was developed in order to address a “gap” at KS3: whereas Speaking and Listening is assessed at GCSE, there is no formal assessment of this earlier. All the English teachers felt there were considerable benefits to this both in assessing individual students and in the effects on the curriculum “balance” more generally: “[I have] a better idea of students for every element, a better understanding of the student’s speaking and listening” (English teacher, June 2007). This had particular benefits for lower attaining students: “[Speaking and Listening] has benefited weaker students and given them confidence” (English teacher, June 2007) and has enabled teachers to “say more [to parents] about the weaker students” (English teacher, June 2007). Teachers themselves benefited from this: “I enjoyed writing schemes of work that focused solely on speaking and listening” (English teacher, June 2007). However, assessing speaking and listening proved problematic. The National Curriculum levels proved inadequate for assessment: they were “too vague” and “[do not] allow you to look at different styles (English teacher, June 2007). As a result, the teachers used modified GCSE criteria as an alternative, although these were still unclear and were more appropriate for more sophisticated thinking. There was a student motivational issue and the teachers felt the students perceived speaking and listening as less important than writing. The ephemeral nature of classroom talk made assessment both difficult and time-consuming. One teacher commented, “assessing reading through small group discussion was useful but getting accurate individual assessments from this is hard and there is a behaviour management issue” (English teacher, June 2007). Providing evidence for standardisation and moderation was difficult. The teachers used video evidence, but this was time-consuming to obtain, moderate and edit. In the event only one school produced videos and watching these was time-consuming. The school with the strongest “culture” of teaching both speaking and listening and drama was most successful at organising and assessing the speaking and listening tasks. Whilst all the teachers felt there were benefits from assessing speaking and listening, only the teachers from the school with prior experience definitely intended to “do more of this thinking next year [to] look at whether [the tasks] enable and recognise speaking and listening” (English teacher, June 2007).

Reading (En2)

Task: A piece also used to assess writing.

Assessment: Reading through student written work using National Curriculum Levels.

Commentary: All the English teachers had difficulty assessing both reading and writing through one combined task. The teachers highlighted a number of inter-related problems related to the particular task itself (a literary criticism piece), to the tension between the criteria for reading and for writing, to doing both within a restricted time frame or whether reading can be assessed through writing, e.g., “Doing it [the literary criticism task] is easy, but assessing both [reading and writing] simultaneously is difficult. It is tempting just to assess writing” (English teacher, June 2007). As with Speaking and Listening further difficulty related to student perceptions of work in English: “students equate writing with work” (English teacher, June 2007). One school introduced a SATS style reading test into Year 8 as “a good test of retrieval” to balance the focus of “extended writing tasks [on] inference” (English teacher, June 2007). Another approach was to assess reading through small group discussion “but getting accurate individual assessments from this is hard and there is a behaviour management issue” (English teacher, June 2007).

Writing (En3)

Tasks: A variety of tasks including non-fiction, fiction, literary criticism and persuasive writing, one piece assessing both reading and writing.

Assessment: National Curriculum levels with some modifications drawn from GCSE criteria.

Commentary: The assessment of writing was the most successful of all the assessment tasks by the English teachers. In carrying out the task, the teachers drew on their past experience of assessing writing particularly, but not exclusively, at GCSE. Nevertheless, the project had a considerable effect on assessment practices. Firstly, the assessment process impacted on the teachers' pedagogy. In particular, the teachers used formative assessment practices more consistently, e.g., peer assessment. Secondly, the teachers became much more aware of what was being assessed and more explicitly negotiating the criteria with students. As a result, students were much more aware of the objectives for a piece of writing, how it would be assessed and they began to gain "insight into layers of meaning and recognising forms and being able to use them" (English teacher, May 2005). "[The students] are clearer about what they are being assessed on" (English teacher, June 2007). This increased emphasis on the quality of interaction and teaching better matched to the needs of students was balanced by the controlled conditions piece which "is a good way of seeing what [the students] can do on their own" (English teacher, June 2007). However, the project highlighted problems and difficulties in the assessment of writing. The portfolio highlighted weaker aspects of curriculum and pedagogy, "what we don't teach very well, like writing to inform" (English teacher, June 2007). As with both reading and speaking and listening, student motivation was a factor and some topics have "more obvious ... like arguing for more pocket money" (English teacher, June 2007). One teacher thought that this, coupled with "poor teaching", means that it is more likely that students will "underperform" in these topics (English teacher, June 2007). The teachers also found the National Curriculum levels vague, inconsistent and lacking in progression, whilst the alternative, GCSE coursework criteria, are geared to more advanced understanding than is generally evident at Year 8.

The specific tasks for each school**School A:**

Speaking and Listening (En1):

- Oral response to "The Edge". Groupwork task involving some peer assessment. (Also used to assess Reading En2)
- Task on "Much Ado About Nothing"

Reading (En2):

- A comparative literary criticism essay. Compare two sonnets by Elizabeth Barratt Browning and by William Shakespeare

Writing (En3):

- Original writing. Following the study of the book, "The Edge", the students wrote a descriptive piece in response to the title, "An unwelcoming place".
- Persuasive writing under timed conditions. Students take the role of outraged residents persuading the people in the locality to get on board their campaign to fight the local leisure centre. This was the Longer Writing Task from the 2006 SATs paper.

School B:

Speaking and Listening (En1):

- Individual extended contribution: “Syson Speech”.
- Drama-focused role-play activity “Court Case”.

Reading (En2):

- Oral response response in a group discussion. From the book, “Holes” the students discussed whether on not Camp Green Lake was a positive experience for Stanley.
- Literary Criticism essay. Response to “Skellig” by David Almond: “How does the character of Skellig develop and change throughout the novel?”
- Persuasive writing under controlled conditions. Read 2 articles for/against T.V. Write a letter to the editor of the newspaper responding to articles and arguing the case for/against. (Also used to assess Writing En3)

Writing (En3):

- Non-fiction writing. History of English leaflet marked for writing to inform
- Fiction writing. “Falling” The student produced the descriptive piece as a result of reading the story of Icarus

School C:

Speaking and Listening (En1):

- Explore how Shakespeare uses staging devices in order to create tension in Act 3 scene 1 of Romeo and Juliet
- Drama piece based on “Private Peaceful”
- Courtroom role-play activity based on “The edge”.

Reading (En2):

- Literary Criticism essay. Compare how two poets develop protest in their poems, examining the use of poetic techniques.
- Persuasive writing under controlled conditions: write an editorial on racism for “The Edge”. (Also used to assess Writing En3)

Writing (En3):

- Non-fiction writing. Travel writing: write an excerpt for “The Holiday Programme” giving information on the destination.
- Fiction writing: write a science fiction story following the conventions of the genre.

APPENDIX D: MATHEMATICS ASSESSMENT TASKS

The Final Assessment

In mathematics, the final assessment used in 2006/7 consisted of two elements: three tasks that were common across all three schools together with three further “non-common” tasks in each school. This was intended to provide both, comparability across schools and to allow for some local flexibility within schools on the basis of school schemes of work, teacher interests etc. A total of six tasks was felt to be manageable for the departments within schools to organise, although, in the event, in-school moderation proved very time-consuming and the within school moderation was largely carried out just by the two KOSAP mathematics teachers in each school.

The Three Common Tasks

Spacers

Task: How many different (cross-, T- and L-shaped) spacers are there on square and rectangular shaped grids of tiles? [For a $n \times m$ grid, these are $(n-1) \times (m-1)$, $2(n-1) + 2(m-1)$ and 4, respectively.] This task was very similar to GCSE algebra coursework.

Assessment: Via students written coursework using amended GCSE coursework criteria, modified to reflect National Curriculum levels.

Commentary: This was the most successful of the assessment tasks used, both in terms of producing evidence of pupils’ work and agreement between teachers’ judgements / gradings, although the teachers had some difficulty aligning the GCSE criteria to National Curriculum. The mathematics teachers had considerable experience of both setting and assessing coursework of this type. It was also the only task used during the Trialling Phase that was not dropped from the final common tasks. Nevertheless, there were difficulties with this task. For example, the teachers had difficulty deciding how to assess work where there had been teacher interventions or advice. In the final task, they drew up procedures specifying and standardising the interventions (and introduction) by the teacher. In addition, teachers in one school felt other teachers in their department lacked both the confidence and the expertise to assess this task reliably. Teachers from one other school felt the marking criteria to be too proscriptive, a criticism they also made of GCSE coursework marking criteria.

Paper Sizes

Task: Compare, describe and classify standard paper sizes (A0-A5) in terms of 1D and 2D ratios. This was a modified GAIM task.

Assessment: Via students’ written work, criteria based on GAIM criteria with considerable simplification and linked to National Curriculum levels.

Commentary: This was a replacement task for Plans and Elevations (see above.) The GAIM task was more acceptable to the teacher who had problems with the original visualisation task. Like the Visualisation task, this highlighted some mismatches between students who could “do” the task in class, but whose write-up did not reflect this. One difficulty with the task was that many students had great difficulty writing up their work, because they had only limited experience of writing up mathematical activities of this kind. In addition, no pupils had commented explicitly about the contrasting relationships between the papers sizes (i.e. the relationships between adjacent paper sizes are the same: two sheets of A4 cover an A3 sheet exactly, the A4:A3 length ratio is $1:\sqrt{2}$ (or 1.4), the A4:A3 area ratio is 1:2). This suggests that student’s mathematical experience of this type of task was rather limited. The simplified assessment criteria caused some tension between the teachers from different schools with other teachers arguing that the criteria left the task too open-ended and to subject to teacher judgement. The task was more successful in the school that developed the criteria than in the other two schools, suggesting that teachers need to reconstruct tasks for themselves.

Newspapers

Task: Compare different newspapers (e.g., tabloid, broadsheet, local) by generating and investigating a hypothesis using simple descriptive statistics.

Assessment: Via students' written work, very simple criteria linked to National Curriculum levels for data-handling.

Commentary: This task was developed by the teachers from one school in order to provide more preparation for the GCSE statistics coursework during KS3. The initial reaction by all the other teachers was very negative. None of these other teachers perceived statistics as an interesting task and all associated it with GCSE coursework. During the development period, it became clear that Statistics coursework was to be abandoned at GCSE. This led to the teachers trying to "open out" the task and focus the task on generating hypotheses, then using means, medians, modes and ranges to compare different newspapers. They re-introduced the task as a hands-on investigation for the teachers. All the teachers viewed this changed focus positively. However, whilst the teachers themselves could generate the task, the actual work from students was poor. Many students had difficulty writing anything, the students' hypotheses were very similar and, whilst almost all students could calculate all the statistics, there was limited evidence that students could interpret the statistical comparisons. As for the other two common tasks, there were some tensions concerning the criteria, which were found by the other teachers to be of limited use. As a result, student work was assessed directly against the National Curriculum.

Additional "Non-common" tasks

School A

1. Painted shapes (Area and Perimeter-type investigation)
Assessment Criteria: GCSE coursework format aligned to National Curriculum levels
2. Coordinates and Graphs. End of unit test. Standard test items devised by school.
Assessment Criteria: Key concepts from scheme of work
3. Data-handling investigation.
Assessment Criteria: Adapted GCSE criteria linked to National Curriculum

School B

1. Plans and elevations (See above.)
2. Hollow shapes (Area and Perimeter-type investigation)
Assessment Criteria: GCSE coursework format aligned to National Curriculum levels
3. Easter summative tests. Standard tests items developed by school.
Assessment Criteria: Breakdown of performance across topics.

School C

1. Poster on Fractions (See above.)
2. Straight line graphs. End of unit test. Standard test items devised by school. Assessment Criteria: Key concepts from scheme of work.
3. School summative tests. Test developed using Testbase software, so all items previously used on KS3 National Tests. Assessment Criteria: National curriculum level based on total score and breakdown of performance across topics.

Mathematics Assessment Tasks Piloted during the Trialling Phase

Plans and Elevations

Task: Construct 3D shapes from multilink given a set of cross-sectional and projection views. Sets of views included “impossible” 3D shapes and views with multiple answers. Students then asked to generate similar problems for each other.

Assessment: Via student written work (pupils’ write-up of the task) and class presentation using National Curriculum levels:

Ma AT3: Shape, Space and Measures, e.g.,

Level 4: Pupils make 3-D mathematical models by linking given faces or edges;

Level 5: When constructing models and when drawing or using shapes, pupils ... use language associated with angle. ... [they] identify all the symmetries of 2-D shapes; Level 6: Pupils recognise and use common 2-D representations of 3-D objects;

Level 7: [Pupils] calculate lengths, areas and volumes in plane shapes and right prisms.

Commentary: This task originated in a class activity. The teachers’ intention was to assess a neglected area of the mathematics curriculum (3D geometry, visualisation) as well as students’ communication and use of mathematics. Initial trials were relatively successful in that they produced assessments of pupils relating both to AT3 and AT1 and highlighted mismatches between student’s investigative and presentational work and between oral and written work as well as geometric and visualisation attainment. However, the teachers found that students in general had great difficulty in the write-up both in terms of writing and producing diagrams. In addition, the description of 3D geometry in National Curriculum levels is restricted. Thus, assessing tasks involved considerable inference and interpretation, placing additional demands on the teachers’ own visualisation skills. The task was dropped as a common task from the final assessment because one teacher from another school had great difficulty with these visualisation skills. It remained as one of the school’s non-common tasks.

Classwork

Task: Two of the schools attempted to assess student classwork within a unit of work. One school assessed a unit on Shape and Space, the other on Algebra.

Assessment: Via student written work in exercise books together with teachers’ ongoing formative feedback and assessed against the aspects of the curriculum highlighted in each schools’ scheme of work.

Commentary: These tasks were used during the Trialling Phase, but not for the final assessments. They were developed partly in response to the teachers’ desire to “look over all students’ shoulders all the time” (Meeting, October 2005) and partly in attempt to align the assessment more closely with classwork. However, when the teachers came to assess the work, they found relatively little differences between the students’ work. In both cases, class work turned out to be a record of what was taught rather than what was learnt. This poor discrimination was exacerbated by schools’ schemes of work. Guided by textbooks and the KS3 Framework, these schemes of work outlined concepts within a much narrower range of National Curriculum (or cognitive) levels than the range of student understanding. One teacher commented on this as follows: “What is the point of what we do? It is as if we are leading them through a forest and just pointing out the trees. It becomes more about how many trees they’ve passed.” (Meeting, June 2006).

Pupil presentations

Task: Two of the schools used presentations by students to assess their understanding on a unit of work. In one case, students were asked to work in pairs to show what they knew about Pythagoras’ Theorem in a poster followed by a class presentation. In the other case, pupils were asked to work individually to annotate a poster to show the links between fractions, decimals and percentages.

Assessment: Via (written) poster and, in the case of the Pythagoras task, oral presentations. Assessed for relevant content / conceptual understanding and for communication skills (as per National Curriculum levels.)

Commentary: These presentational tasks were developed by teachers in response to their consideration of what it means to be “good at maths”. The fractions activity was used for part of the final assessment as one

of the school's non-common tasks. The Pythagoras poster was used during the trialling phase in Spring 2006, but not for the final assessments. For the Pythagoras task, the teachers envisaged the oral presentation to be one of the most important aspects for assessment, but found the ephemeral nature of classroom talk difficult to capture and the assessment, discrimination and evidencing of student talk difficult. In order to make the task manageable, the teachers used a simple tick list, but the other teachers found this evidence unconvincing. For both tasks, there was a tension between structuring the task in order that the students could (and would) demonstrate their knowledge, whilst leaving the task sufficiently open that students could decide for themselves what were the most important aspects to present. Both cases highlight a validity issue. In both cases, this was the only occasion when students were asked to make formal presentations in mathematics. Without more frequent opportunities for the teaching and learning of mathematical talk and communication, it is unlikely that either students or teachers will develop their skills in this area.

APPENDIX E: THE PILOT PROJECT

The pilot study, funded by the DfES, ran from January 2005 - March 2006 and focused on an exploration of existing summative practices of the 12 English and mathematics teachers and development and trialling of new or adapted assessment activities. The work of the pilot study encouraged the English and mathematics teachers to be more reflective about their current assessment practices and to consider the potential of summative assessment activities used by the other two schools and by the other subject domain in the project. This has led the teachers to question their current practice, to hone and refine aspects of their current practice and to propose some changes to their procedures. Our interest for the main study lay in the processes and practices that the teachers drew on and developed to manage to do this, both in their own practice and across their departments. We were interested in the effect this has had on their summative practices overall and, in particular how confident and rigorous they were in deciding the final teacher assessment component of any summative system. A further interest was the effects these changes had on their formative practice and the ways they attempted to resolve the summative-formative interface.

APPENDIX F: PROJECT PERSONNEL

King's College, London:

Christine Harrison	(project manager)
Jeremy Hodgen	(mathematics specialist)
Bethan Marshall	(English specialist)
Paul Black	(consultant)
Natasha Serret	(part-time research officer).

Oxfordshire LA:

Dorothy Kavanagh	(assessment adviser)
Christine Dickinson	(Key Stage 3 consultant)
Jan Evans	(Key Stage 3 consultant)
Cath Borien	(Key Stage 3 consultant)