



Working Paper 2

**Summative assessment by teachers:
evidence from research and its
implications for policy and practice**

Assessment Systems for the Future Working Paper 2, draft 4

Summative assessment by teachers: evidence from research and its implications

Contents

1. Introduction
2. Purposes and used of assessment
3. Pros and cons of summative assessment by teachers
 - 3.1 Arguments in favour
 - 3.2 Arguments against
4. Research evidence
 - 4.1 Research reviews
 - 4.2 Review procedures
 - 4.3 Definition of terms
5. Findings: the reliability and validity of summative assessment by teachers
 - 5.1 Variables that affect dependability
 - 5.2 The role of assessment criteria
 - 5.3 Addressing potential bias
 - 5.4 The conditions affecting dependability
6. Findings: the impact of teachers' summative assessment on students, teachers and the curriculum
 - 6.1 Impact on students
 - 6.2 Impact on teachers and the curriculum
 - 6.3 Conditions affecting the nature and extent of the impact
7. Implications
 - 7.1 Implications for national local assessment policy
 - 7.2 Implications for school management
 - 7.3 Implications for teachers
 - 7.4 Implications for teachers education and professional development
 - 7.5 Implications for researchers

List of studies used in the two reviews

1. Introduction

This paper summarises the research evidence revealed by the latest in a series of reviews of research that have explored what can be learned from research studies of the uses of assessment in education. Earlier reviews dealt with several aspects of assessment and its uses, including assessment and classroom learning¹, the impact of summative assessment and tests on students' motivation for learning² and the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills³.

One of the implications arising from the review of research on the impact of testing on students' motivation for learning was that some of the negative effects of tests could be avoided by making greater use of summative assessment by teachers. Other arguments in favour, and against, the use of teachers' judgements for summing up the attainments of their students prompted the search for research evidence, in two further reviews, that are the basis for this Working Paper.

Before listing those arguments we begin here by setting summative assessment by teachers in the context of the purposes and uses of assessment. We then summarise the research evidence revealed by two reviews of research – one focusing on the reliability and validity of summative assessment by teachers and the other on the impact on students and teachers of teachers' summative assessment. The final section proposes implications of the findings for educational practice, teacher professional development, assessment policy, and research.

2. Purposes and uses of assessment

It is now commonplace to distinguish among different purposes of assessment, such as formative (or assessment for learning), summative (or assessment of learning), monitoring, evaluation and research. Figure 1 sets the purpose that is the focus of interest here in the context of different purposes, agents and uses of assessment.

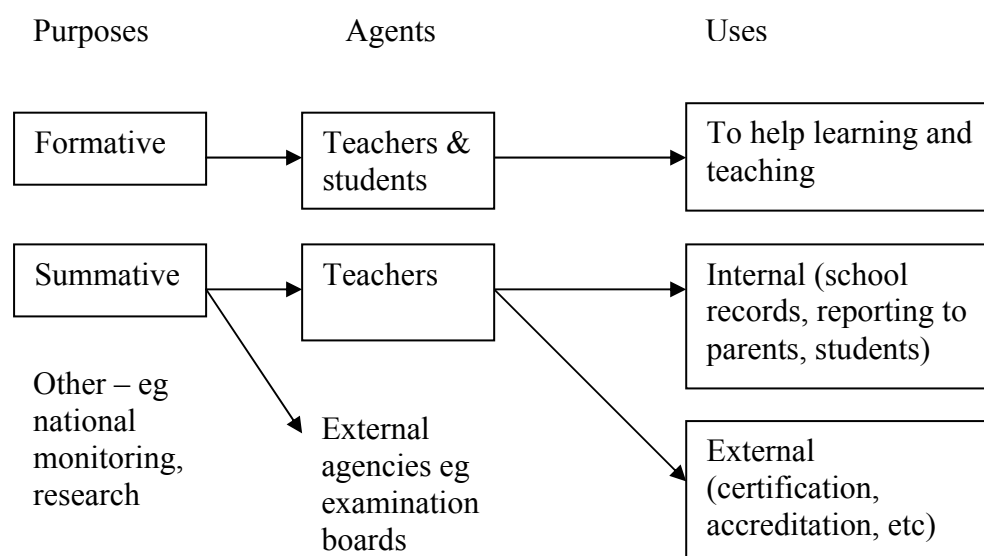
Formative assessment has the single purpose of informing learning and teaching, whilst summative assessment has the purpose of reporting on learning achieved at a certain time. Summative assessment, however, has more than one use, for there is a variety of ways in which the information about student achievement at a certain time can be used. These can be grouped into two main uses – 'internal' and 'external' to the school community. Internal uses include using regular grading for class and school records, keeping track of student progress, informing decisions about courses to follow where there are options within the school, reporting to parents and to the students themselves. External uses include certification by examination bodies or for vocational qualifications, selection for the next stage of education or for employment, monitoring the school's performance and school accountability.

¹ The review by Paul Black and Dylan Wiliam was published in full in *Assessment in Education*, 1998, 5 (1) and summarised in *Inside the Black Box*, 1998, giving rise to several subsequent short 'Black Box' publications, now published by nferNelson.

² The review findings are summarised in the ARG pamphlet *Testing Motivation and Learning* (2002), available on the ARG website www.assessment-reform-group.org.uk The full review report is available on the EPPI-Centre website http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_one.htm

³ http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_two.htm

Figure 1



The concern here is mainly with the impact of assessment for external uses, since generally there is not an issue about the role of teachers in regular summative assessment for internal school records and reporting to parents and to students. It is assessment for external uses that acquires 'high stakes', meaning uses that are associated with the status of the school and in some cases directly with its financial support or with the salaries of individual teachers. When stakes are high, summative assessment can acquire a stranglehold on what is taught and how it is taught. However, internal uses are not entirely free from high stakes, for parents can exert pressure where they are not satisfied with results, and results for classes or departments can be set against each other during internal school evaluation. Research shows that where external assessment has high stakes the effect is for internal assessment to emulate external tests and examinations, increasing the impact on students (37)⁴.

3. Pros and cons of summative assessment by teachers

3.1 Arguments in favour

Arguments in favour of a move towards greater use of assessment by teachers for external uses have been advanced for some time, the main points being that

- As part of their regular work, teachers can build up a picture of students' attainment across the full ranges of activities and goals. This gives a broader and fuller account of achievement than can be obtained through tests, which can only include a restricted range of items.
- There is less pressure on students and teachers compared with external tests and examinations; freedom from test anxiety means that the assessment is a more valid indication of students' achievement.
- There can be greater freedom for teachers to pursue learning goals in ways best suited to their students, rather than being constrained by what is perceived as necessary in order for students to pass tests.

⁴ Numbers in brackets throughout this Working Paper refer to the research studies listed at the end of the paper.

- There is the potential for information about students' on-going achievements to be used formatively, to help learning, as well as for summative purposes.
- Assessment by teachers can facilitate a more open and collaborative approach to summative assessment in which students can share in the process through self-assessment and derive a sense of progress towards 'learning goals' as distinct from 'performance goals'.

3.2 Arguments against

However there are equally strongly advocated arguments against assessment by teachers having a significant role in summative assessment:

- There is a widespread assumption, supported by research evidence, of unreliability and bias in teachers' assessment.
- Being responsible for summative assessment would bring with it an additional workload for schools and teachers.
- Although all teachers are necessarily involved in summative assessment for 'internal' purposes (eg school records, grouping and setting, reporting to parents, students, other teachers), involvement in summative assessment for 'external purposes' (eg certification, selection, school accountability) emphasises the dual role that is required as teacher and assessor.
- Over-elaborate moderation procedures for quality assurance could constrain the operation of teachers' summative assessment so that only 'safe' and routine approaches are used.
- Resources are required for procedures such as moderation, assessment planning and professional development to ensure the necessary dependability of assessment for external uses.

4. Research evidence

4.1 Research reviews

What does research say about these opposing claims? We report here the findings from two systematic reviews of research on summative assessment, focused on these issues. The research questions were, for the first:

- What is the research evidence of the reliability and validity of assessment by teachers for the purposes of summative assessment?
- What conditions affect the reliability and validity of teachers' summative assessment?

And for the second:

- What is the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes?
- What conditions and contexts affect the nature and extent of the impact of using teachers' assessment for summative purposes?

In both reviews the research evidence was used to address the further question of the implications of the findings for policy and practice in summative assessment.

4.2 Review procedures

Both systematic reviews were conducted using the procedures and tools of the EPPI-Centre. This involved a wide-ranging search for research studies, written in English,

of assessment for summative purposes in schools for pupils between the ages of 4 and 18, which reported evidence relevant to the research questions. The search for studies involved scanning relevant electronic databases and journals online, following up citations in other reviews, hand-searching journals held in the library, and using personal contacts. Successive rounds of applying explicit criteria resulted in the identification of the most relevant studies, which were analysed in depth⁵. This ensured that the synthesis and any conclusions were based on the best evidence available. Judgements were made as to the strength of evidence relevant to the review provided by each study. In the synthesis, greater weight was given to studies providing the strongest evidence. All judgments in applying criteria, data extraction and evidence weight were made by two people working independently and afterwards resolving differences by further reference to the evidence in the study.

The number and type of studies found in these procedures were as follows:

Review of the reliability and validity of summative assessment by teachers	Review of the impact on students and teachers of summative assessment by teachers
<p>The initial search for studies in this review identified 431 papers. These were reduced in the successive stages of the review process to 30 for in-depth analysis and data extraction. The studies are listed at the end of this publication; reference to the evidence they provide is indicated by numbers in this summary of findings.</p>	<p>343 relevant studies were found in the initial search. After the various stages of screening, 23 studies were included in the in-depth review. Four of the studies included in the review of reliability and validity were also found to provide relevant data for this review. Again they are referenced by number from the list given at the end of the paper.</p>
<p>Eleven studies involved primary students (aged 10 or below) only, 13 involved secondary students (aged 11 or above) only and six were concerned with both primary and secondary students.</p>	<p>Eleven of the 23 studies involved primary school students (aged 10 or below) only, six involved secondary students (aged 11 or above) only and five were concerned with both primary and secondary students.</p>
<p>Eighteen studies were classified as involving assessment of work as part of, or embedded in, regular activities. Three were classified as portfolios, two as projects and nine were either set externally or set by the teacher to external criteria. In the vast majority, students were assessed by teachers using external criteria.</p>	<p>Twenty studies were classified as involving assessment of work as part of, or embedded in, regular activities. Three were classified as portfolios, two as projects and eight were either set externally or set by the teacher to external criteria.</p>
<p>The most common use of the assessment in the studies was for national or state-wide assessment programmes, with six studies relating to certification and another six to informing parents (in combination with other uses).</p>	<p>The most common use of the assessment in the studies was for internal school purposes, with four studies related to assessment for certification and another three to external purposes that had high stakes for the school.</p>

Sections 5 and 6 report separately the findings from each of the reviews. The implications in section 7 draw on findings from both reviews.

⁵ Using the *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research*, Version 0.9.7.

4.3 Definition of terms

The definition of summative assessment by teachers (see box) adopted in the review focused on assessment by teachers, using their professional judgment, of their own students. It does not, therefore, include the role of teachers in setting tests for others to use or marking tests given to students other than their own.

Summative assessment by teachers:

The process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgement, and to report at a particular time on their students' achievements.

Reliability:

How accurate the assessment is.

Validity:

How well what is assessed matches what it is intended to assess.

Dependability:

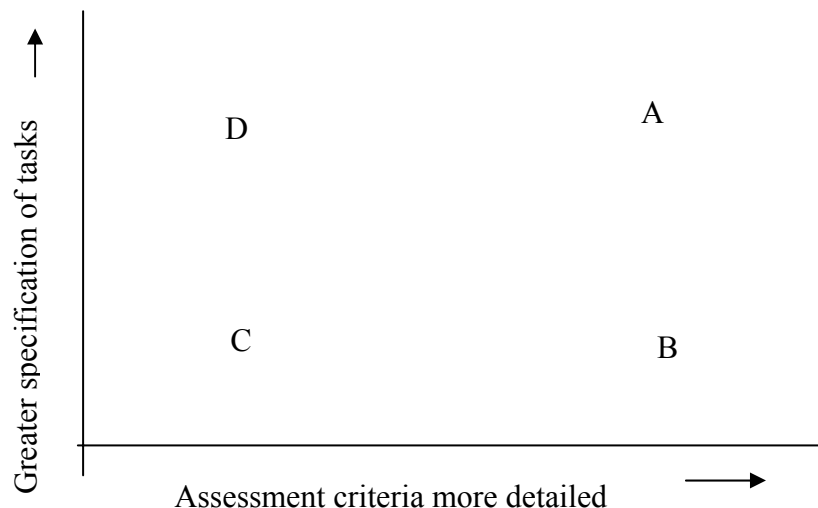
The extent to which reliability is optimised whilst ensuring validity.

5. Findings: the reliability and validity of summative assessment by teachers

5.1 Variables that affect dependability

The assessment practices in the research studies that were reviewed varied in many ways, but most significantly in the two features: the extent to which the tasks assessed are specified and the extent to which the criteria applied in judging performance are specified. These key dimensions can be used to identify the major characteristics of the more and less reliable and valid approaches. The variation in the specification of the task or tasks can be envisaged as spread along a dimension (y-axis) from unspecified, when the assessment is based on the whole range of regular work, to tight specification passing through a mid point where types of tasks to be included in the assessment may be specified. For each type of task there are also different approaches spread along another dimension, from loosely specified criteria to closely specified criteria for judgement. This 'criterion' dimension (x-axis) extends from general judgements, grading or ratings where no precise meaning is given to the labels used, to detailed criteria which match particular tasks, passing through a mid point where brief descriptions are used to define points on a grading or rating scale. Figure 2 indicates four main types of approaches that are defined by the intersection of these dimensions.

Figure 2



When tasks are unspecified, tight criteria can guide the selection of work assessed (area B), whilst general, non-specific criteria leave the validity of the sample of work in the hands of the teacher (area C). When tasks are closely specified (D and A), the validity depends on the selection of tasks made in designing the assessment programme and on how well the criteria match the specified tasks.

The extent to which tasks for assessment by teachers are specified is at the heart of the reasons for including assessment by teachers in assessment systems; the more tightly specified the tasks the less opportunity there is for a broad range of learning outcomes to be included in the assessment. Approaches to assessment in area A, where both task and criteria are closely specified, as, for instance, in an externally devised practical science investigation or oral assessment in a foreign language, can provide reliable data (16, 19, 43). In these cases, the teacher is acting as an administrator of an instrument devised by others, and indeed is administering an external test. This meets the definition of summative teachers' assessment noted above only because the application of the criteria requires professional knowledge and could not be carried out by someone without this knowledge and who is on the spot.

Not surprisingly, the dependability of approaches where neither task nor criteria are well specified has been found to be low (30, 41). But even where tasks, or particular pieces of work, are specified but the criteria are not defined, the reliability is low (30, 41). This happens, for instance, in portfolio assessment where types of work to be included are specified and teachers rate the work (such as on a five point scale where the only guidance is 5= high, 1= low).

5.2 The role of assessment criteria

Greater dependability is found in area B, where there are detailed, but generic, criteria that allow evidence to be gathered from the full range of classroom work. The subject profile approach (40) is an example here. Sets of criteria are provided relating to achievement in various subjects and in specific aspects within subjects. Rather than trying to match a particular piece of work with a particular criterion, teachers take evidence from several relevant pieces and form a judgment on the basis of the best match between the evidence and the criteria. The criteria also serve the additional function of focusing attention on the outcomes of particular kinds of work, so that teachers are alerted to looking for particular behaviours and are less likely to miss them. Other studies (16, 24, 43) also suggest that when criteria are well specified, teachers are able to make reliable judgments.

The assessment by teachers in the National Curriculum Assessment in England and Wales allows evidence to be used from the regular classroom work. There is evidence that how teachers go about this varies (18, 23, 38), but this does not in itself necessarily affect the reliability. Teachers vary in their teaching approaches and any less variation in assessment practice would not be expected. Certainly, variation according to the nature of the subject and how it is taught is to be expected if assessment is truly embedded in regular work.

But the validity of approaches that leave tasks and the sampling of the domain unspecified depends on the extent to which the evidence actually gathered is a good sample of work in the areas concerned. While having detailed criteria leads teachers to consider work from relevant areas, if these areas are not well covered in the implemented curriculum, the opportunity for assessment is clearly limited. There is evidence that consistency in applying criteria seems to depend upon teachers being clear about the goals of the work (24, 30) and on the thoroughness with which

relevant areas of the curriculum are covered in teaching (11). The context of the school's support and value system has also been identified as having a role in how assessment by teachers is practised (21). Conditions associated with greater dependability include the extent to which teachers share interpretations of criteria and develop a common language for describing and assessing students' work.

5.3 Addressing potential bias

There was evidence of bias and error in teachers' summative assessment in the findings of some studies. This was generally due to teachers taking into account information about non-relevant aspects of students' behaviour (2), or being apparently influenced by gender, special educational needs, or the general or verbal ability of a student in judging performance in a particular task (5, 6, 13, 14, 31, 39, 44, 46, 49). Several researchers claim that bias in teachers' assessment is susceptible to correction through focused workshop training (19, 31) although this review did not specifically include studies of the impact of such training. However, participation of teachers in developing criteria was found to be an effective way of enabling the reliable use of the emerging criteria (16, 24, 40).

5.4 The conditions affecting dependability

It was possible to identify from the research evidence conditions that affect the dependability of the assessment. The main ones were:

- Detailed criteria describing levels of progress in various aspects of achievement enable teachers to assess students reliably on the basis of regular classroom work (16, 33, 40).
- It is important for teachers to follow agreed procedures if assessment by teachers is to be sufficiently dependable to serve summative purposes (21, 23, 30, 41).
- The training required for teachers to improve the reliability of their assessment should involve teachers as far as possible in the process of identifying criteria so as to develop ownership of them and understanding of the language used. Training should also focus on the sources of potential bias that have been revealed by research (16, 24, 30, 31, 40, 41).
- Dependable assessment needs protected time for teachers to meet for training and moderation of their judgments and to take advantage of the support that others can give (19, 21, 38).
- Moderation through professional collaboration is of benefit to teaching and learning as well as to assessment (19, 21, 38).

6. Findings: the impact of teachers' summative assessment on students, teachers and the curriculum

6.1 Impact on students

The focus of this review was the impact of the *process* of summative teacher assessment, not the impact of the outcome of the assessment (for instance the reaction of a student to a high or low score). The separation into impact on students and impact on teachers is artificial since, in the context of teachers conducting the assessment, any impact on students is mediated by impact on teachers. However, studies reporting students' reactions to the assessment process and teachers' perceptions of impact on learning could be distinguished from those reporting impact on teachers and their views of impact on teaching and the curriculum.

Where the assessment took the form of assessment of coursework for an external award, a positive impact on students was reported (7). Students found the

coursework motivating because it provided them with an element of choice and the incentive to acquire and use new skills in finding things out for themselves and in communicating. The students were, however, aware that it was the product and not the process of the work that counted. They were less aware than their teachers of the constraints of the assessment criteria that made teachers reluctant to allow students to take control of the coursework. In another approach, where a single examination at the age of 16 was replaced by a series of graded tests taken throughout the secondary school (28), students also responded positively to the scheme and almost all preferred it to a single end of course examination. Exceptions to general approval were found among the lower achieving students for whom the frequent tests were a constant reminder of their failure to progress as quickly as others. In both of these studies researchers found that students had a much poorer grasp of the criteria and aims of the work than the teachers assumed. They recommended more effort to share these with students through providing examples and models of what was required. This need was confirmed in other studies (6, 45).

Assessment for external purposes based on teachers' judgements across a range of student work, rather than on specified tasks, was associated with a strong positive impact on teaching and learning when it was built into teachers' planning, not added on to satisfy official requirements (21, 23). The introduction of teachers' assessment related to levels of the National Curriculum in England and Wales was perceived by teachers as having a positive impact on students' learning (23). The positive impact was further enhanced when teachers worked collaboratively towards a shared understanding of the goals of the assessment and of procedures to meet these goals (21).

When assessment by teachers was for internal purposes, there was evidence that extending teachers' assessment practices beyond looking at products to include learning processes and students' explanations led to better student learning. In this context, the nature of the feedback given was found to be an important factor in determining students' effort in further tasks. Effort was motivated by non-judgemental feedback that gave information about how to improve (4, 8). How teachers present classroom assessment activities was found to be a factor in affecting whether students sought to achieve goals of learning or goals of achieving high marks or other rewards (4). Also, using grades as rewards and punishment was found to encourage extrinsic motivation (28, 36).

6.2 Impact on teachers and the curriculum

Teachers vary in their response to the requirements of assessment for external summative purposes, some keeping rigidly to the regulations and others being prepared to interpret them in the best interests of their students, without stepping outside the intentions of the procedures (34, 50). There was evidence that their response was influenced by the high stakes of the assessment (50). Teachers were adversely affected by the requirements of conducting external summative assessment when it was perceived as taking up too much time from teaching (1, 3). There was, however, compensation for the time spent in the value that they gained about their students (1, 3, 30) and about learning opportunities for students that needed to be extended (47, 48).

The assessment by teachers for internal purposes, when unguided by agreed criteria, could be influenced by non-achievement factors, such as students' behaviour, effort, and attendance, as already noted (2, 10). Reports of such assessment provided little dependable information to those receiving them (10, 36). The information is more useful to others where teachers are able to internalise the nature of progression(26). The existence of criteria, and particularly involvement in

identifying them, helps teachers in understanding the meaning of learning outcomes (47). But there is evidence of a need to distinguish between externally devised checklists, which encourage a mechanistic approach to assessment, and the use of criteria that identify qualitative differences in progression towards learning goals. Close external control of teachers' summative assessment was considered to inhibit teachers gaining detailed knowledge of their students (29).

6.3 Conditions affecting the nature and extent of the impact

The research evidence pointed to conditions that affect the impact of teachers' summative assessment. These were that:

- New assessment practices are likely to have a positive impact on teaching if teachers find them of value in helping them to learn more about their students and to develop their understanding of curriculum goals. Time to experience and develop some ownership of assessment practices enhances their positive impact (1, 3, 15; 17, 30).
- When high stakes judgments are associated with teachers' assessment, one effect is for teachers to reduce assessment tasks to routine events and restrict students' opportunities for learning from them. The existence of high stakes also encourages some teachers to give high grades where there is doubt, which may not be in the students' interests (7, 21, 34, 50).
- The use of shared criteria for assessing specific aspects of achievement leads to positive impact on students and on teaching; in the absence of such guidance there is little positive impact of teachers' summative assessment on teaching and a potential negative impact on students (2, 10, 23, 32, 36).
- The process that teachers use in setting assessment tasks and in grading, impacts on students' motivation for learning, particularly their goal orientation, when grades are used as rewards or punishments. The negative impact can be alleviated by ensuring that students have a firm understanding of assessment processes and of criteria (4, 7, 28, 45).
- Summative assessment by teachers has a more positive impact on teachers and teaching when integrated into practice than when concentrated at certain occasions (2, 7, 8, 23, 28, 29, 30, 32, 48).
- Opportunities for teachers to share and develop their understanding of assessment procedures enable them to review their teaching practice, their view of students' learning and their understanding of subject goals. Such opportunities have to be sustained over time and preferably should include provision for teachers to work collaboratively across, as well as within, schools (15, 17, 21, 23, 25, 47).

7. Implications

Evidence from reviews of research such as these, whilst drawing on the specific contexts reported in the studies reviewed, can offer pointers as to how policy and practice in summative assessment by teachers might best be developed. In the following sections we identify some implications for those concerned with assessment policy in national and local government, for school management teams, for classroom teachers, for those involved in the initial education and continuing professional development of teachers, and for researchers.

It is evident from the above summaries that there are common themes in the findings from the two reviews. Many of the conditions that support greater dependability of teachers' summative assessment are also ones that facilitate a positive impact on students and teachers. But there are also clear messages about the changes in

practice that are required, and the support that teachers need in making these changes, if teachers' summative assessment is not only to become dependable but to have a role in helping teaching and learning.

7.1 Implications for national and local assessment policy

Policies concerning assessment practices in schools are typically formulated either at national or regional level by the government departments and agencies that establish the frameworks within which schools are required to operate. The interpretation and implementation of such policies by schools and by individual teachers are likely to ensure that the experience of students and teachers will be variable from one classroom to another. However, it remains the case that national and regional level policies determine the policy environment within which schools and teachers make their decisions about assessment practices. Some implications of the reviews for those responsible for assessment policy are that:

- It is important to consider the use of assessment in deciding the strengths and weaknesses of using teachers' assessment in a particular case. For instance, when assessment is fully under the control of the school and is used for informing pupils and parents of progress ('internal' purposes), the need to combine the judgements of teachers with other evidence (eg tests) may be less than when the assessment results are used for 'external' purposes, such as accountability or the school or selection or certification of students.
- The short-comings of external examinations and tests as well as those of assessment by teachers need to be borne in mind in deciding the balance between them for external summative purposes
- There needs to be greater recognition of the difference between uses of summative assessment and of how to match the way such assessment conducted with a particular use.
- Using teachers' assessment for summative purposes can support valid assessment of key learning processes as well as assessment of learning outcomes related to higher level cognitive skills and has the potential for positive effects on students and on teachers, without the negative effects associated with external tests and examinations.
- Summative assessment by teachers has most benefit when teachers use evidence gathered over a period of time, and with appropriate flexibility in choice of tasks, rather than from an event that takes place at a particular time. This enables information to be used formatively to inform learning and teaching as well as summatively.
- It is important to provide professional development for teachers in undertaking assessment that addresses the potential sources of bias and error in teachers' summative assessment. However, the process of moderation should also be recognised as a means of developing teachers' understanding of learning goals and related assessment criteria as well as a means of increasing the dependability of the outcome.
- When changes are made in assessment practices, time must be allowed for schools to assimilate unfamiliar procedures summative assessment into their practice and to design appropriate classroom programmes. Imposing unduly tight regulation on schools inhibits the beneficial impact that summative assessment by teachers can have on students and teaching.
- Using the results of student assessment for high stakes school accountability reduces the validity of the assessment, whether this is conducted by teachers or by external tests and examinations.
- There is a need for resources to be put into identifying detailed criteria that are linked to learning goals, not specially devised assessment tasks. This will

support teachers' understanding of the learning goals and may make it possible to equate the curriculum with assessment tasks.

7.2 Implications for school management

School management teams have to respond appropriately to government initiatives and when new ones appear they will take attention and time away from previous ones. It is necessary, therefore to ensure that new practices are well established in school policies and practice if their value is to be maintained. Evidence from the reviews of teachers' summative assessment showed that the positive impact on teachers and teaching found initially was severely reduced as attention turned to other developments. Individual teachers were left to take responsibility rather than responsibility being treated seen as a whole school matter. .

Some implications from the evidence in the reviews are that senior management should

- Include all teachers, not just those involved in assessment for external purposes, in regular school meetings where assessment is planned, issues discussed and the school policy kept under review.
- Ensure commitment to the school's assessment policy by involving teachers in developing it, rather than imposing it 'top down'.
- Ensure teachers have protected time for moderation of their judgements of students' work.
- Ensure that responsibility for internal moderation procedures is clearly assigned.
- Set up effective procedures, as appropriate, for monitoring and improving the quality of teachers' assessment that is used for external purposes. This will include monitoring the processes and the moderation of outcomes.

7.3 Implications for teachers

The benefit to teaching and learning of teachers undertaking summative assessment of their own students is increased when teachers:

- Are clear about the goals of learning and have internalised the progression in skills and understanding they aim to help students develop. In this way they can interpret student performance in terms of progression rather than using a checklist of specific and unconnected behaviours. Summative assessment can thus help teachers' understanding of learning goals as well as facilitating more detailed knowledge of their students.
- Help students to understand the criteria by which their work is assessed. This is likely to mean providing and discussing examples that illustrate work that meets the criteria.
- Make explicit to all concerned – colleagues, parents and students - the basis of the marks and grades they assign for internal school purposes. Marks or grades given for achievement in learning should not be influenced by non-academic factors, such as behaviour and participation, which should be reported separately as appropriate.
- Emphasise learning processes and outcomes and not the attainment of a high grade when presenting assessment tasks to students. This avoids the encouragement of extrinsic motivation, which leads to shallow learning.
- Ensure that all learning goals, both processes and outcomes, are assessed.
- Are aware of the possible sources of bias in their assessments, including the 'halo' effect, and follow procedures that guard against such bias.
- Ensure that evidence is gathered from performance in meaningful tasks rather than relying on itemised checklists which record students' isolated aspects of achievement rather than development towards learning goals.

- Take part in discussing assessment constructively and positively within the school.

7.4 Implications for teacher education and professional development

Initial teacher education and profession development should reflect the significant role of assessment in education by giving appropriate time and attention to it. This should include attention to assessment for both formative and summative purposes, and the circumstances in which information can serve both purposes. In relation to summative assessment, programmes of initial teacher education and professional development should:

- Enable trainees, teachers and classroom assistants to recognise the different uses of summative assessment and to become aware of the ways of conducting it that are suited to these uses.
- Give teachers and trainees experience of developing criteria that indicate progression so that teachers and classroom assistants understand the meaning and role of such criteria developed by others.
- Give teachers and trainees opportunity to practice different ways of gathering information and procedures for ensuring shared understanding and application of assessment criteria for summative uses.
- Support school management in establishing a school culture in which assessment is seen as integral to, and having a positive impact, on teaching and learning.

7.5 Implications for researchers

The reviews have revealed several gaps in our knowledge about the processes by which teachers arrive at judgements of students' achievement and thus how dependability might be increased. Given the recent growth of interest in the UK in the use of teachers' summative assessment, at all stages of schooling, it is matter of urgency to improve the evidence base. For example, there need to be more studies of:

- How teachers go about assessment for different purposes, what evidence they use, how they interpret it, etc. This should include investigation of the reasons for the difference between teachers' estimates of performance as compared with those of moderators so that appropriate action can be taken.
- How teachers perceive the dual role as teacher and assessor and how this affect students.
- Ways of establishing the dependability of teachers' summative assessment need to be developed. The essential and important differences between teachers' assessment and tests should be recognised by ceasing to judge the dependability of teachers' summative assessment in terms of how well it agrees with test scores.
- Factors that support teachers use of summative assessment to improve students learning experience; that is, whether, and if so, how the formative use of assessment can be integrated with the summative use.
- Direct comparison of different approaches used by teachers in summative assessment to investigate whether they make any difference to outcomes or to impact on students.
- The role of student self-assessment in summative assessment and the impact of developing students' awareness of assessment criteria and of providing exemplification of learning goals.
- What changes to accountability procedures would preserve the integrity of teachers' assessment and minimise pressures to give inflated grades or levels.

List of studies used in the two reviews

1. Abbott, D., Broadfoot, P., Croll, P., Osborn, M. and Pollard, A. (1994) Some sink, some float: national curriculum assessment and accountability, *British Educational Research Journal*, 20, 155 - 174.
2. Bennett, R. E., Gottesman, R. L., Rock, D., A. and Cerullo, F. (1993) Influence of behaviour perceptions and gender on teachers' judgments of students' academic skill, *Journal of Educational Psychology*, 85, 347-356.
3. Bennett, S. N., Wragg, E. C. et al. (1992). A longitudinal study of primary teachers' perceived competence in, and concerns about, National Curriculum implementation. *Research Papers in Education* 7(1): 53 - 78.
4. Brookhart, S. M. and DeVoge, J. G. (1999). Testing a theory about the role of classroom assessment in students motivation and achievement. *Applied Measurement in Education* 12(4): 409 - 425.
5. Brown, C. R. (1998) An evaluation of two different methods of assessing independent investigations in an operational pre-university level examination in biology in England, *Studies in Educational Evaluation*, 24, 87-98.
6. Brown, C. R., Moor, J. L., Silkstone, B. E. and Botton, C. (1996) The construct validity and context dependency of teacher assessment of practical skills in some pre-university level science examinations, *Assessment in Education*, 3, 377 - 391.
7. Bullock, K., Bishop, K., et al. (2002). Learning from coursework in English and geography. *Cambridge Journal of Education* 32(3): 325 - 340.
8. Carter, C. R. (1997). Assessment: Shifting the Responsibility. *Journal of Secondary Gifted Education* 9(2): 68-75.
9. Chen, M. and Ehrenberg, T. (1993) Test scores, homework, aspirations and teachers' grades, *Studies in Educational Evaluation*, 19, 403 - 419.
10. Cizek, G. J., Fitzgerald, S. M. et al. (1995/1996). Teachers' Assessment Practices: Preparation, isolation and the kitchen sink. *Educational Assessment* 3(2): 159 - 179.
11. Coladarci, T. (1986) Accuracy of teachers' judgments of students responses to standardized test items, *Journal of Educational Psychology*, 78, 141 - 146.
12. Crawford, L., Tindal, G. and Steiber, S. (2001) Using oral reading rate to predict student performance on statewide achievement tests, *Educational Assessment*, 7, 303 - 323.
13. Delap, M. R. (1995) Teachers' estimates of candidates' performance in public examinations., *Assessment in Education*, 2, 75-92.
14. Delap, M. R. (1994) An investigation into the accuracy of A-level predicted grades, *Educational Research*, 26, 135-149.
15. Flexer, R., Cumbo, K., Borko, H., Mayfield, V. and Marion, S. F. (1995). How 'messing about' with performance assessment in mathematics affects what happens in classrooms. *CSE Technical Report 396* Los Angeles, CRESST: 49.
16. Frederiksen, J. and White, B. (2004), Designing assessment for instruction and accountability: an application of validity theory to assessing scientific inquiry. In (Ed) Wilson, M. *Towards Coherence between Classroom Assessment and Accountability*, 103rd Yearbook of the National Society for the Study of Education part II. Chicago: National Society for the Study of Education. 74-104
17. Gipps, C. and Clarke S. (1998). *Monitoring consistency in teacher assessment and the impact of SCAA's guidance materials at Key Stages 1, 2, and 3. Final Report*. London, Institute of Education: 87.
18. Gipps, C., McCallum, B. and Brown, M. (1996) Models of teacher assessment among primary school teachers in England, *The Curriculum Journal*, 7, 167 - 183.
19. Good, F. J. (1988) Differences in marks awarded as a result of moderation: some findings from a teachers assessed oral examination in French, *Educational Review*, 40, 319 - 331.
20. Good, F. J. and Cresswell, M. (1988) Can teachers enter candidates appropriately for examinations involving differentiated papers? *Educational Studies*, 14, 289-297.
21. Hall, K. and Harding, A. (2002) Level descriptions and teacher assessment in England: towards a community of assessment practice, *Educational Research*, 44, 1.
22. Hall, K. and Harding, A. (1999). Teacher Assessment of Seven Year Olds in England: A Study of its Summative Function. *Early Years* 20(1): 19 - 28.
23. Hall, K., Webber, B., Varley, S., Young, V. and Dorman, P. (1997) A study of teacher assessment at Key Stage 1, *Cambridge Journal of Education*, 27, 107 - 122.
24. Hargreaves, D. J., Galton, M. J. and Robinson, S. (1996) Teachers' assessments of primary children's classroom work in the creative arts, *Educational Research*, 38, 199 - 211.
25. Hiebert, E. and Davinroy, K. (1993). Dilemmas and issues in implementing classroom-based assessment for literacy. *CSE Technical Report 365* Los Angeles, CRESST: 25.
26. Hill, M. (2002). Focussing the Teacher's Gaze: Primary teachers reconstructing assessment in self managing schools. *Educational Research for Policy and Practice* 1(12): 113 - 125.
27. Hopkins, K. D., George, C. A. and Williams, D.D. (1985) The concurrent validity of standardized achievement tests by content area using teachers' ratings as criteria, *Journal of Educational Measurement*, 22, 177-82.

28. Iredale, C. (1990). Pupils' attitudes towards GASP (Graded Assessments in Science Project). *School Science Review* 72(258): 133-137.
29. Johnston, P. H., Afflerbach, P. and Weiss, P. B. (1993). Teachers' assessment of the teaching and learning of literacy. *Educational Assessment* 1(2): 91 - 117.
30. Koretz, D., Stecher, B. M., Klein, S. P. and McCaffrey, D. (1994) The Vermont Portfolio Assessment Program: findings and implications, *Educational Measurement: Issues & Practice*, 13, 5 - 16.
31. Levine, M. G., Haus, G. J and Cort, D. (1987) The accuracy of teacher judgment of the oral proficiency of high school foreign language students, *Foreign Language Annals*, 20, 45-50.
32. McCallum, B., McAlister, S., Brown, M., and Gipps, C. (1993). Teacher assessment at Key Stage One. *Research Papers in Education* 8(3): 305 - 328.
33. Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y. and Atkins-Burnett, S. (2001) Trusting teachers' judgments: a validity study of a curriculum-embedded performance assessment in kindergarten to Grade 3, *American Educational Research Journal*, 38, 73-95.
34. Morgan, C. (1996). The teacher as examiner: the case of mathematics coursework. *Assessment in Education* 3(3): 353-376.
35. Papas, G and Psacharopoulos, G. (1993) Student selection for higher education: the relationship between internal and external marks, *Studies in Educational Evaluation*, 19, 397 - 402
36. Pilcher, J.K. (1994). The value-driven meaning of grades. *Educational Assessment* 2(1): 69-88.
37. Pollard, A., Triggs, P., Broadfoot, P., McNess, E. and Osborn, M. (2000) *What Pupils Say: Changing Policy and Practice in Primary Education*. London: Continuum. Chapters 7 and 10)
38. Radnor, H. A. (1996) *Evaluation of Key Stage 3 Assessment Arrangements for 1995. Final Report* University of Exeter, Exeter, pp. 181.
39. Reeves, D. J., Boyle, W. F. and Christie, T. (2001) The relationship between teacher assessment and pupil attainments in standard test/tasks at key stage 2, 1996 - 8, *British Educational Research Journal*, 27, 141 - 160.
40. Rowe, K. J. and Hill, P. W. (1996) Assessing, recording and reporting students' educational progress: the case for 'subject profiles', *Assessment in Education*, 3, 309-352.
41. Shapley, K. S. and Bush, M. J. (1999) Developing a valid and reliable portfolio assessment in the primary grades: building on practical experience, *Applied Measurement in Education*, 12, 11-32.
42. Sharpley, C. F. and Edgar, E. (1986) Teachers' ratings vs standardized tests: an empirical investigation of agreement between two indices of achievement, *Psychology in the Schools*, 23, 106 - 111.
43. Shavelson, R. J., Baxter, G. P. and Pine, J. (1992) Performance assessments: political rhetoric and measurement reality, *Educational Researcher*, 21, 22 - 27.
44. Shorrocks, D. Daniels, S., Staintone, R. and Ring, K (1993) Testing and Assessing 6 and 7 year olds. *The evaluation of the 1992 Key Stage 1 National Curriculum Assessment*. National Union of Teachers and Leeds University School of Education
45. Stables, A. (1992). Speaking and listening at key stage 3: some problems of teacher assessment. *Educational Research* 34(2).
46. Thomas, S., Madaus, G. F., Raczek, A. E. and Smees, R. (1998) Comparing teacher assessment and the standard task results in England: the relationship between pupil characteristics and attainment, *Assessment in Education*, 5, 213 - 246.
47. Valencia, S. W. and Au, K. H. (1997). Portfolios across educational contexts: issues for evaluation, teacher development and system validity. *Educational Assessment* 4(1): 1 - 35.
48. Whetton, C., Sainsbury, M. Hopkins, S., Ashby, J., Christophers, U., Clarke, J., Heath, M., Jones, G., Punchers, J., Schagen, I., Wilson, J. (1991). *A Report on Teacher Assessment*. London, SEAC.
49. Wilson, J. and Wright, C. R. (1993) The predictive validity of student self-evaluations, teachers' assessments, and grades for performance on the verbal reasoning and numerical ability scales of the differential aptitude test for a sample of secondary school students attending rural Appalachia schools, *Educational & Psychological Measurement*, 53, 259-70.
50. Yung, H-W, B. (2002). Same assessment, different practice; professional consciousness as a determinant of teachers; practice in a school-based assessment scheme *Assessment in Education* 9(1): 97 - 117.