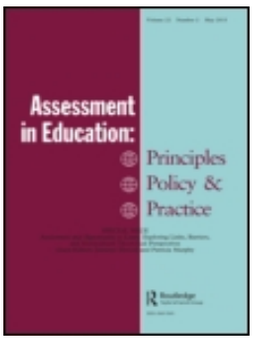

Routledge
Taylor & Francis Group

Assessment in Education: Principles, Policy & Practice

ISSN: 0969-594X (Print) 1465-329X (Online) Journal homepage: http://www.tandfonline.com/loi/caie20

Assessment and Learning: differences and relationships between formative and summative assessment

Wynne Harlen & Mary James

To cite this article: Wynne Harlen & Mary James (1997) Assessment and Learning: differences and relationships between formative and summative assessment, Assessment in Education: Principles, Policy & Practice, 4:3, 365-379, DOI: 10.1080/0969594970040304

To link to this article: http://dx.doi.org/10.1080/0969594970040304

Assessment and Learning: differences and relationships between formative and summative assessment

WYNNE HARLEN[1] & MARY JAMES[2]
[1]Scottish Council for Research in Education, 15 St. John Street, Edinburgh, EH8 8JR, UK & [2]University of Cambridge School of Education, Shaftesbury Road, Cambridge CB2 2BX, UK

ABSTRACT *The central argument of this paper is that the formative and summative purposes of assessment have become confused in practice and that as a consequence assessment fails to have a truly formative role in learning. The importance of this role is argued particularly in relation to learning with understanding (deep learning). It is pointed out that the requirements of assessment for formative and summative purposes differ in several dimensions, including reliability, the reference base of judgements and the focus of the information used. This challenges the assumption that summative judgements can be formed by simple summation of formative ones. An alternative procedure for linking formative and summative assessment is proposed such that their separate functions are preserved.*

Introduction

The aim of this paper is to share concerns that arise from the particular approaches to assessment adopted in the countries of the UK, but which will almost certainly have resonance in other countries, particularly in Europe, the USA, Australia and New Zealand. In outline the problems can be stated as follows:

- that formative and summative assessment are both included in national assessment policies and in theory have different roles, but the way in which they have been related to each other in official documents means that the essential differences between them have been smothered;
- a consequence of the conflation of summative and formative purposes may be that either there is little genuine formative assessment (or what there is may not be recognised as such) or that teachers are struggling to meet both requirements and experiencing assessment overload;
- because formative assessment has to be carried out by teachers, there is an assumption that all assessment by teachers is formative, adding to the blurring of the distinction between formative and summative purposes and to teachers changing their own on-going assessment into a series of 'mini' assessments each of which is essentially summative in character;

0969-594X/97/030365-15

- the equating of formative assessment with teachers' assessment coupled with the effective down-grading (in England, at least) of teachers' judgements in comparison with externally devised tests or tasks on account of the latter being used to create league tables, has led to the neglect of support for formative assessment;
- there is a need to recognise in theory and in practice the differences in function and characteristics between formative and summative assessment and to find a way of relating them together that preserves their different functions; in particular we want to argue that it is not necessary, and indeed it is not helpful, to be concerned with strict criterion-referencing in formative assessment.

These problems have become more prominent in the UK since the introduction of the educational reforms of the late 1980s. The descriptions of attainment in the National Curriculum have provided criteria (at levels 1-10, recently revised to 1-8, in the National Curriculum and levels A-E in the Scottish Guidelines) for teachers to use in their assessment and have thus added rigour that was formerly missing. However, emphasis in the guidance to teachers has been on the application of the criteria to pupils' achievements, for the purpose of deciding what level they have reached, to the neglect of the genuinely formative use of assessment.

The framework of assessment and testing in the National Curriculum for England and Wales was set by the report of Task Group on Assessment and Testing, known as TGAT, published in 1988 by the Department of Education and Science (DES) and the Welsh Office (WO) (DES/WO, 1988). The recommendations in this report, which included the framework of levels of attainment, were accepted instantly and entirely by the government without consultation. However, during the course of the implementation of the report's recommendations there have been significant changes which have been well documented by Daugherty (1995) and Black (1997), and as a result not all present practice in assessment in England can be linked back to this report. Nevertheless, it was the TGAT report that put the terms *formative, diagnostic, summative* and *evaluative* into common circulation and defined them. The distinction between formative and summative was made mainly in terms of purpose and timing:

> —formative, so that the positive achievements of a pupil may be recognised and discussed and the appropriate next steps may be planned
> —summative, for the recording of the overall achievement of a pupil in a systematic way. (DES/WO, 1988, para. 23)

The assumption that these were not different *in kind* is clear in the claim that some purposes could be served by combining assessment originally made for different purposes: 'It is possible to build up a comprehensive picture of the overall achievements of a pupil by aggregating, in a structured way, the separate results of a set of assessments designed to serve formative purposes' (DES/WO, 1988, para. 25).

The message that formative and summative are easily related in this way and are of the same kind was reinforced by concern for standardisation: 'in order to ensure comparability' (DES/WO, 1988, para. 44). This concern is rightly raised in relation to summative assessments which may be used to make comparisons between pupils

or to provide results which can be aggregated to give whole-class or whole-school profiles. However, set beside the above view that summative assessments can be formed by simple aggregation of formative ones, it leads to the inevitable conclusion that formative assessments must also be 'standardised'.

We believe that this relationship between formative and summative assessment is simplistic and has brought about a fundamental confusion in teachers' minds about these two kinds of assessment. The reliability of summative assessment has suffered from the confusion. For example, evidence from the evaluation of implementation of national assessment and testing in Scotland found that it was quite common for teachers to share this view of how a summative assessment of a child's work is reached: 'What would be a piece of C work for one child might not be for another, it depends on their background' (Harlen *et at,* 1995). Similar findings were reported by Gipps *et al.* (1995). However, the focus of this paper is on the detrimental effect of this confusion of purposes on formative assessment and on the role that assessment has to play in teaching for understanding. A central purpose of this paper is to propose a different relationship which will preserve the essential function of formative assessment in learning.

In the first part of the paper we consider the kind of learning we are concerned to bring about. This is followed by revisiting the nature of formative assessment and its role in learning, leading to a comparison of the characteristics of formative and summative assessment. In later sections we consider a different approach to linking formative and summative assessment and the consequences for developing this approach.

Learning with Understanding

We begin from the assumption that an important aim of education is to bring about learning with understanding. This has been called 'learning as an *interpretative* process' (Säljö, 1979; Broudy, 1988) or 'deep learning' (Entwistle & Ramsden, 1983; Marton *et al,* 1984; Entwistle & Entwistle, 1991). The term 'real' learning extends the notion of learning with understanding to suggest that it involves interaction with people, ideas, things and events in the real world. These are not the only kinds of learning. There are some things that are probably most efficiently learned by rote, such as number bonds, spellings, multiplication tables; these are skills that are most useful to us when they have been practised sufficiently to become automatic. There are also bodies of knowledge (facts and information) that it is reasonable for society to expect teachers to teach and for pupils to learn, although there are still debates about *what* knowledge and *whose* knowledge is most important. However, the exponential increase in the amount of factual information in recent years and for the foreseeable future, coupled with the rapid changes in the nature of employment, indicate that there should be far greater emphasis on learning which can be transformed and applied to new circumstances than on learning facts and procedures applicable only in situations closely similar to those in which they were learned. Indeed, it is the need of the whole population to be able to translate and interpret what they learn that makes the case for universal schooling (Broudy,

1988). From his review of relevant research, Crooks concludes that 'there seems to be a strong case for encouraging the development of deep strategies from the early years of the educational system' (Crooks, 1988, p. 447).

Crooks refers here to the simple but powerful way of identifying the approaches to learning which lead, on the one hand to understanding and on the other to rote memorisation, in terms of the distinction between *deep* learning and *surface* learning, defined as follows (see, for example, Ausubel *et al.*, 1978; Marton *et al.*, 1984; Entwistle & Entwistle, 1991):

Deep Learning Approach	*Surface Learning Approach*
An intention to develop personal understanding	An intention to be able to reproduce content as required
Active interaction with the content, particularly in relating new ideas to previous knowledge and experience	Passive acceptance of ideas and information
Linking ideas together using integrating principles	Lack of recognition of guiding principles or patterns
Relating evidence to conclusions	Focusing learning on assessment requirements

Between surface learning and deep learning Marton *et al.* (1984) also proposed an intervening category which they called 'strategic learning'. This reminds us that efficient learning is often a combination of both surface and deep learning, for if we were to learn everything in depth we would have time to learn very little. Likewise, if everything was surface learning we could hardly describe ourselves as educated at all. Assessment has a role in all kinds of learning. In memorising facts and learning physical skills it is used to find out what facts or skills have been acquired and the feedback it provides to help further learning is in terms of what has not been learned. Assessment has quite a different role in learning with understanding and it is this that is the concern in this paper and which is now considered.

When something is learned with understanding (deep learning, 'real' learning) it is actively understood and internalised by the learner. It makes sense in terms of a learner's experience of the world and is not simply a collection of isolated facts which have been memorised. As noted above, it differs from rote learning essentially in that it is linked to previous experience and so can be used in situations different from that in which it was learned. Contemporary cognitive psychology supports the notion that understanding involves creating links in the mind and that 'making sense' of something depends on these links. Isolated pieces of information do not have links to existing mental frameworks and so are not easily retained in the mind. The identification and creation of links to existing frameworks depends on the active participation of the learner and on the familiarity of the context of the material to be learned. Understanding, in this view, is the process of construction and reconstruction of knowledge by the learner. What is known and understood will, of course, change with new experience and as new ideas and skills are presented to help make sense of it. Thus the characteristics of this learning are that it:

- is progressively developed in terms of big ideas, skills for living and learning, attitudes and values;
- is constructed on the basis of previous ideas and skills;
- can be applied in contexts other than those in which it was learned;
- is owned by the learner in the sense that it becomes a fundamental part of the way he or she understands the world; it is not simply ephemeral knowledge that may be memorised for recall in examinations but subsequently forgotten.

It follows that to promote this kind of learning, what is needed are learning experiences that:

- are well matched to the existing point of development of the ideas, skills, attitudes and values;
- have continuity with, and build on, previous experience;
- relate to current interests and experience;
- are perceived by learners as relevant, important, stimulating and valued for themselves, rather than simply for their usefulness in passing tests and examinations.

The provision of learning experiences with these kinds of characteristics depends on the teacher:

- having a thorough and deep understanding themselves of the subject matter to be taught, how pupils are likely to learn it and the difficulties and misunderstandings they are likely to encounter;
- having a clear idea of the progression in the ideas, skills, etc. which are the goals of learning and the course pupils are likely to take in this development;
- being able to recognise the point in this development reached by their pupils;
- knowing and being able to use various strategies to find out and to develop pupils' ideas, skills, etc.

Our focus here is on the last two items in this list. Knowing about pupils' existing ideas and skills, and recognising the point reached in development and the necessary next steps to take, constitutes what we understand to be *formative assessment.* This is consistent with Sadler's (1989) definition. However, Sadler goes further to argue that, if improvement in learning is to take place, *students* need to come to hold a concept of quality roughly similar to that held by the teacher. They also need to be able to monitor the quality of what is being produced during the act of production, and to draw on a range of strategies to close any gap between their actual performance and the standard they are aiming for. This implies that part of the teacher's role in assessment for learning is to help students to come to these metacognitive understandings. Formative assessment, therefore, is essentially feedback (Ramaprasad, 1983), both to the teacher and to the pupil about present understanding and skill development in order to determine the way forward. Assessment for this purpose is part of teaching; learning with understanding depends on it. To use information about present achievements in this way means that the

progression in ideas and skills must be in the teacher's mind—and as far as possible in the pupils'—so that the next appropriate steps can be considered.

Summative assessment has a quite different purpose, which is to describe learning achieved at a certain time for the purposes of reporting to parents, other teachers, the pupils themselves and, in summary form, to other interested parties such as school governors or school boards. It has an important role in the overall educational progress of pupils but not in day-to-day teaching as does formative assessment. As the next section of this paper attempts to show, it is the distinction between, and the articulation of assessment for these two purposes that is central to using assessment to improve educational standards.

For either purpose, if understanding is to be assessed, methods are required that involve learners in *using* their knowledge and linking it to real contexts. It cannot be assessed by asking for the recall of isolated, decontextualised pieces of information. The straightforward reproduction of knowledge rather than its application favours rote learning, and assessment which demands no more than this will inevitably shift teaching and learning away from understanding towards the memorisation of the information necessary to succeed in the assessment.

The Nature of Formative Assessment

Formative assessment has to be carried out by the teacher, but so is a great deal of assessment for summative purposes (including certification in cases where course work is part of the process). Thus in this section we discuss the characteristics of formative assessment, particularly in relation to how these differ from the characteristics of summative assessment.

Unlike summative assessments, which may be either criterion-referenced or norm-referenced, formative assessments are always made in relation to where pupils are in their learning in terms of specific content or skills. To this extent, formative assessment is, by definition, *criterion-referenced.* At the same time, it may also be *pupil-referenced* (or *ipsative).* This means that a judgement of a pupil's work or progress takes into account such things as the effort put in, the particular context of the pupil's work and the progress that the pupil has made over time. In consequence, the judgement of a piece of work, and what is fed back to the pupil, will depend on the pupil and not just on the relevant criteria. The justification for this is that the individual circumstances must be taken into account if the assessment is to help learning and to encourage the learner. If formative assessment were purely criterion-referenced it would be profoundly discouraging for many pupils who are constantly being faced with failure. This hybrid of criterion-referenced and ipsative assessment does not matter as long as this information is used *diagnostically* in relation to each pupil, which is consistent with the notion that formative assessment is essentially part of teaching.

The claim that criterion-referenced systems often only thinly disguise norm-referenced systems would lead to the contentious notion that there is some degree of norm-referencing in formative assessment. It is true that any attempt to articulate a trajectory of development of knowledge, skill and understanding in any subject

domain often implies assumptions about 'normal' stages of development and progression. Also, judgements about an individual's progress in relationship to others are sometimes helpful in identifying whether there is an obvious problem that needs to be tackled urgently. This is probably a main reason why parents continue to be so concerned about where their child is in relation to the attainments of others of the same age. The point to be made here in the context of formative assessment, however, is that whilst norm-referenced assessment might help teachers recognise the existence of a problem, it can offer them no help in knowing what to do about it and may simply have a deleterious effect by labelling or pigeon-holing pupils. In order to contribute to learning through teaching, assessments need to reveal the specific nature of any problems; this can only be achieved by a combination of criterion- and pupil-referenced assessments.

Essentially this kind of formative assessment, which involves using information about their learning gathered from observing pupils, listening to them discussing informally with their peers as well as when talking to the teacher, reviewing written work and other products, and using their self-assessments, has always been part of teachers' work. However, it has often been carried out less systematically than is required to serve its purpose effectively and not always used in helping to identify the next steps in learning. As noted above, concern for deeper learning and the constructivist view of learning have done much to draw attention to aspects of pupils' thinking that teachers should take into account. Gaining access to the ideas and mental frameworks that pupils have already formed, accurate or otherwise, is an integral part of teaching for understanding and requires teachers to adopt new strategies for lesson planning and different forms of questioning (Harlen, 1996).

It is important to recognise that the reality of formative assessment is that it is bound to be incomplete, since even the best plans for observing activities or setting certain tasks can be torpedoed by unanticipated events. Moreover, the information will often seem contradictory. Students are always changing and may appear to be able to do something in one situation but not in another. Such evidence is a problem where the purpose is to make a judgement about whether a pupil fits one category, criterion, or one level or another. However, where the purpose is to inform teaching and help learning, the fact that a pupil can do something in one context but apparently not in another is a positive advantage, since it gives clues to the conditions which seem to favour better performance and thus can be a basis for taking action. In this way the *validity* and usefulness of formative assessment is demonstrated and enhanced. Validity is vitally important to formative assessment because it cannot claim to be formative unless it demonstrably leads to action for improved learning; hence relevance to the goals of learning is paramount.

However, it is not necessary to be over-concerned with *reliability* in formative assessment since the information is used to inform teaching in the situations in which it is gathered. Thus there is always quick feedback for the teacher, who usually has opportunities to use observations of the response to one intervention as information in making the next one. Pryor & Torrance (1996, p. 214) give examples of this process in action. Through this rapid loop of feedback and adjustment between teacher and learner, the information inevitably acquires greater reliability.

This is not to say that teachers do not need any help with this important part of their work, but the help required is to be found in how to identify significant aspects of pupils' work and to recognise what they mean for promoting progress.

In the feedback between teacher and pupils, both parties need to be involved in decisions about the next steps in learning. The learner who recognises for him or herself how to move forward is likely to take responsibility for making the necessary effort. Involving pupils in their own assessment means that they must know what are the aims of their learning. Communicating these aims is not easy, but the rewards of successfully attempting it are quite considerable, not only for help in assessment, but also in the obvious potential for self-direction in learning. Direct communication of complex learning objectives and criteria of achievement is unlikely to be successful, but pupils can develop understanding of them through experiences designed to involve them in looking critically at their own work. In the primary school these might involve asking pupils to select their best pieces of work and then discussing why the particular ones were chosen. More generally, the comments made by the teacher on pieces of work convey what is expected and valued. Teachers of older pupils can more explicitly share with the pupils the criteria they use both in assessing practical skills and marking written work.

Differences between Formative and Summative Assessment

In summary, we have attempted to distinguish formative assessment from summative assessment by suggesting that the characteristics of formative assessment are that:

- it is essentially positive in intent, in that it is directed towards promoting learning; it is therefore part of teaching;
- it takes into account the progress of each individual, the effort put in and other aspects of learning which may be unspecified in the curriculum; in other words, it is not purely criterion-referenced;
- it has to take into account several instances in which certain skills and ideas are used and there will be inconsistencies as well as patterns in behaviour; such inconsistencies would be 'error' in summative evaluation, but in formative evaluation they provide diagnostic information;
- validity and usefulness are paramount in formative assessment and should take precedence over concerns for reliability;
- even more than assessment for other purposes, formative assessment requires that pupils have a central part in it; pupils have to be active in their own learning (teachers cannot learn for them) and unless they come to understand their strengths and weaknesses, and how they might deal with them, they will not make progress.

In contrast, the characteristics of summative assessment are that

- it takes place at certain intervals when achievement has to be reported;
- it relates to progression in learning against public criteria;

- the results for different pupils may be combined for various purposes because they are based on the same criteria;
- it requires methods which are as reliable as possible without endangering validity;
- it involves some quality assurance procedures;
- it should be based on evidence from the full range of performance relevant to the criteria being used.

Linking Formative and Summative Assessment whilst Preserving their Separate Functions

Although earlier we criticised the direct link made between formative and summative assessment, it is not being suggested that information gathered by teachers for formative purposes should not be used when they come to make summative assessments. This would be wasteful and in any case impossible in practice, for teachers cannot ignore knowledge that they have of pupils. Instead, we regard it as essential to distinguish different ways of *arriving at an assessment* for different purposes. But for reliable assessment, as required for summative purposes, there must be certain conditions on the use of this information, specifically that:

(i) it is reviewed strictly against the criteria of what students are expected to achieve at certain ages/stages;
(ii) the criteria are applied holistically, using judgements as to the 'best fit';
(iii) there is some way of ensuring that the judgements of one teacher are comparable with those of other teachers.

(i) Using External Criteria

Our concern in this paper is with criterion-referenced assessment, since this provides information about the abilities and understanding developed by the student. Criteria are used in both formative and summative assessment, enabling the information from the assessment to be used in planning for the student's future learning opportunities; formative in the short-term and summative in the longer-term. Because formative assessment is carried out so that it can be used in helping teaching and learning, the type of information needed is concerned with the learning in the activities as they take place. It will therefore concern the ideas and skills that can be developed in particular activities. We can call these 'small' ideas because they are likely to be specific to the activity. Of course, the teacher will have in mind 'bigger' ideas, of wider application, towards which (s)he intends the pupils to make progress. Several 'small' ideas will have to be linked together to form gradually bigger and bigger ideas. For example, if pupils collect creatures from the garden and attempt to keep them in their classroom, they are learning about the conditions needed for survival by the particular animals that they find. Their teacher would want them to link their ideas about these particular creatures to their experiences with other living things—keeping pets, learning about animals in the wild, etc.—to form general ideas about the needs of living things and eventually to link the

characteristics of animals to the features of the habitats in which they live. The characteristics of living things and adaptation to habitat are 'big' ideas (big because they relate to more than the particular animals studied) that will become bigger still as pupils' experience extends and supports a deeper understanding of these concepts.

Summative assessment is concerned with progress towards the big ideas rather than with the learning in specific activities. The external criteria to be applied invariably relate to these big ideas and in using them the teacher will wish to judge the extent to which the students have shown development towards achieving them by, for example, being able to apply ideas in contexts different from those in which they were learned. Thus (s)he will look across several activities to judge the extent to which there is evidence of the development of the kind of understanding indicated in the criteria to be applied summatively.

(ii) Using Criteria Holistically

At this point it is useful to keep in mind that the kind of information that is gathered by teachers in the course of teaching is not tidy, complete and self-consistent, but fragmentary and often contradictory. The unevenness, as mentioned above, is not a problem but an advantage for formative purposes, helping to indicate what supports or hinders achievement for a particular pupil. However, these uneven peaks and troughs have to be smoothed out in reporting performance for summative purposes. Thus although some of the same *evidence* can be used for formative and summative purposes, for the latter it has to be reviewed and aligned with criteria applied uniformly across all pupils. This means looking across the range of work of a pupil and judging the extent to which the profile as a whole matches the criteria in a holistic way. Where criteria are identified in developmental sequence, this is a matter of deciding the level at which a students' work as a whole best fits the criteria, accepting that not every piece of work will meet the criteria and not every criterion at the 'best fit' level will be met.

(iṅ) Improving Reliability

In the process of applying criteria, professional judgement is inevitably required since criteria cannot be so detailed and specific that their application is purely procedural. The reliability of the judgements can be enhanced through various approaches to quality assurance. Whilst it is not the purpose of this paper to review these approaches (see Harlen, 1994a), it is worth mentioning some frequently used in the context of teachers' assessment for summative purposes. One such approach is for teachers to discuss together their judgements of collections of students' work. Given the points made about using a holistic approach, it is important for these to be portfolios of work and not single pieces. Another approach is to build up a bank of examples, again of portfolios, that can be used as reference points in making judgements. A third way is to provide externally devised tasks or tests that have been

validated for the purpose of indicating certain levels of achievement for teachers to use to check their judgements.

Using Evidence for Formative and Summative Purposes in Practice

In summary, the alternative to using the same *results* of assessment for both purposes is to use relevant *evidence* gathered as part of teaching for formative purposes but to review it, for summative purposes, in relation to the criteria which will be used for all pupils. This means that formative assessment can remain a mixture of criterion-referenced and pupil-referenced assessment, as is required for providing a positive response to pupils and encouraging their learning. At the same time the use of information gathered as part of teaching, appropriate for formative assessment but which could be misleading or even confusing if used directly for summative assessment, is filtered out in the process of reviewing information relevant to the criteria being applied (in the level descriptions for example). In other words, summative assessment should mean summing up *the evidence,* not summing across a series of judgements or completed assessments as implied by TGAT.

What this approach can mean in practice is illustrated by an example taken from a package of material devised to support primary teachers in Scotland in using diagnostic assessment to help pupils learn science (Scottish Council for Research in Education (SCRE), 1995). The material describes in detail the work of a group of 10—11-year-olds investigating camouflage as part of an environmental studies project. They planned the details of the investigation themselves following broad suggestions from the teacher. It involved placing cut-out figures painted in different colours against different backgrounds and judging which was easiest to see from a distance. There were many opportunities, some taken and some missed, to plan and conduct a controlled investigation. Although the pupils worked in groups of six, there was plenty of opportunity for the teacher to observe the performance of all the pupils in relation to different aspects of investigation—planning, finding out, recording, interpreting and reporting (Scottish Office Education Department (SOED), 1993). The account given in the pack provides detailed evidence about two of the pupils in-the form of descriptions of their actions and reproductions of their work. A commentary highlights significant aspects of the work from which are drawn suggestions for further progress. For example, in relation to their planning:

> Ben (who had shown a clear idea of how to carry out the investigation in the way selected) would be helped by being asked to think of different ways of approaching the investigation. What do you want to find out from your investigations? Can you put this in the form of a question you want to answer? In what different ways could you set about answering this questions?
>
> Anna (who had not shown a good grasp of what was being tested in the investigation) would benefit from being challenged to give reasons for her planned actions and to distinguish between things which are important to the investigation and those which are not. Was it important that this was

> done in this way? Why? If someone else were going to do this what would be the most important things to tell them? (SCRE, 1995, p. 23)

The same information is also used to help the teacher to judge the level of the work against the criteria set out in the attainment targets of the curriculum. In this matter, however, the account is accompanied by a warning. The work described was an extended investigation carried out at intervals over several days. It is noted that the 'level' of the pupils' work appeared to change during this time:

> Had an assessment been made of their work in planning the investigation at the end of the first hour or so, most would seem to have been operating at the level of 'suggests possible strategies for carrying out an investigation including resources without much thought for a controlled investigation....' However, as the work proceeded their continuing planning showed attention to some of the controls necessary for the comparison of colours to be fair. This cautions against making assessments at too early a point in an investigation when pupils are considering a number of possibilities in a fairly superficial manner rather than a few with more rigour, as they may do later. (SCRE, 1995, p. 22)

Another example shows how a missed opportunity to distinguish between the formative and summative use of the information leads to a neglect of the formative value of the information. One element of the 'Exemplification of Standards' material distributed to all schools in England and Wales in 1995 by the Schools Curriculum and Assessment Authority (SCAA) and the Advisory Council on Assessment and the Curriculum for Wales (ACAC) is a video and booklet containing evidence of pupils engaged in speaking and listening (English Attainment Target 1) and judged to be at various levels from 1 to 8. The Key Stage Three (lower secondary) material includes footage of a girl named Nicole, for whom English is 'an additional language', who is seen contributing to four different activities. A teacher viewing this video might notice that Nicole watches the faces of peers very closely, although sometimes obliquely, and sometimes angles herself so that she can read the text from which they are reading. She is often to the side of group interaction and has difficulty breaking into a fast verbal exchange. Occasionally her contributions are 'talked over' by others who are more forceful. However, when the activity gives her an opportunity to 'have the floor' she speaks quietly and slowly but more confidently and her contributions are structured and comprehensible. This kind of evidence might be used formatively by the teacher to indicate how Nicole's learning in this area might be extended by building on her listening skills, by acknowledging the tremendous progress she has made in competent use of her second language, by helping her with sentence constructions that she finds especially difficult, by providing her with more opportunities to speak in formal presentations where she cannot be interrupted by more confident peers, and by working with the whole group on their understanding of the nature and dynamics of group discussion to allow better pacing, turn taking, listening, inclusion, etc. But none of this is mentioned in the material accompanying the video because it is 'designed to help teachers make consistent judgements about

which level best describes a pupil's performance'. Thus the commentaries on Nicole's contributions relate strictly to the general criteria embedded in the level descriptions. The peaks and troughs and idiosyncrasies of her performance are ironed out for the purpose of coming to the following summary and overall judgement:

> Although she perhaps lacks confidence, Nicole contributes clearly and positively in discussions. She makes substantive points, gives reasons and is able to argue for her views when challenged. She is beginning to ask questions of others and take account of their views. She adjusts her speaking to more formal situations although she is not fully confident in standard English. Overall, Nicole's performance is best described by Level 5. (SCAA/ACAC, 1995, p. 30)

Many other examples similar to this last one indicate that the fundamental distinction between formative and summative assessment has not been fully articulated. Formative assessment involves, as we have argued, a combination of criterion-referencing and pupil-referencing, whereas summative assessment involves a combination of criterion-referencing and norm-referencing. Formative and summative assessment may relate to each other in that they share a set of common criteria which are agreed expectations in terms of desired outcomes, but beyond this they are essentially different phenomena with different assumptions and different methods. Some of the same evidence may be used for different purposes but it will be used in different ways.

Developing the New Approach

It is essential to provide help for teachers with both formative and summative assessment and in a way which disentangles the two and enables teachers to use assessment in a genuinely formative way to help pupils' learning. This would include guidance on types of feedback from teachers necessary to maintain pupil motivation, as well as on identifying specific aspects of attainment or good performance and what to do to help further improvement (see Tunstall & Gipps, 1996). For formative assessment, all four of the qualities required for fostering learning with understanding (mentioned above) need to be developed within initial teacher training and continuing professional development. In particular, teachers may need assistance in identifying 'next steps' in learning, perhaps more in relation to some subjects, such as science, than in others. The further development of exemplar materials may be important here, but only if the materials are directed towards the need to make 'next steps' decisions, as in the Scottish material, rather than overall summative judgements. Teachers may also value examples of techniques for gaining access to pupils' ideas and for involving them in self-assessment and in deciding their 'next steps'. Such collaboration between teacher and pupils facilitates pupils taking some responsibility for their own learning.

For summative purposes a rather different view of the process of summarising needs to be taken than the ones used to date. Each piece of work, each observation

made by the teacher, is used to build up a picture which is not determined by one or two events, since to demonstrate understanding of a concept or skill it has to be applied in different contexts. It is inappropriate to provide exemplars in terms of single pieces of work which are judged to meet the description at a certain level. Experience shows that this has a damaging washback effect on formative assessment. Rather, exemplar material in the form of portfolios of work from one pupil could be used to help teachers to develop the skills of applying the level descriptions in a holistic manner and to recognise that not every piece of work will fit the descriptions and neither will every element of the description at a level be represented in the portfolio. This may seem rather a loose procedure, but we have repeatedly pointed out (e.g. Harlen, 1994b, p. 139; 1996, p. 147) that assessment is not an exact matter, can never be, and if we try to treat it as such we may damage the very learning we are striving to bring about. It is less a technical matter of measurement and more a human act of judgement, albeit based on sound evidence. In our programmes for initial and in-service teacher training we need to provide opportunities for them to develop confidence and expertise in making and using such judgements about, and for, learning. Then we might have confidence that standards of real learning will rise to meet the demands of the next century.

References

AUSUBEL, D.P., NOVAK, J.S. & HANESIAN, H. (1978) *Educational Psychology: a cognitive view* (New York, Holt, Rinehart & Winston).

BLACK, P. (1997) Whatever happened to TGAT?, in: C. CULLINGFORD (Ed.) *Assessment Versus Evaluation*, pp. 24-50 (London, Cassell).

Broudy, H.S. (1988) *The Uses of Schooling* (New York, Routledge).

CROOKS, T.J. (1988) The impact of classroom evaluation practices on pupils, *Review of Educational Research*, 58, pp. 438-481.

DAUGHERTY, R. (1995) *National Curriculum Assessment: a review of policy 1987-94* (London, Falmer Press).

DES/WO (1988) *National Curriculum Task Group on Assessment and Testing—a report* (London, DES).

ENTWISTLE, N.J. & ENTWISTLE, A.C. (1991) Forms of understanding for degree examinations: the pupil experience and its implications, *Higher Education*, 22, pp. 205-227.

ENTWISTLE, N.J. & RAMSDEN, P. (1983) *Understanding Pupil Learning* (London, Croom Helm).

GIPPS, C., BROWN, M., MCCALLUM, B. & MCALLISTER, S. (1995) *Institution or Evidence?* (Buckingham, Open University Press).

HARLEN, W. (1994a) Towards quality in assessment, in: W. HARLEN (Ed.) *Enhancing Quality in Assessment*, pp. 139-145 (London, Paul Chapman).

HARLEN, W. (Ed.) (1994b) *Enhancing Quality in Assessment* (London, Paul Chapman).

HARLEN, W. (1996) *The Teaching of Science in Primary Schools* (London, David Fulton).

HARLEN, W., MALCOLM, H. & BYRNE, M. (1995) *Assessment and National Testing in Primary Schools* (Edinburgh, SCRE Research Project).

MARTON, F., HOUNSELL, D.J. & ENTWISTLE, N.J. (Eds) (1984) *The Experience of Learning* (Edinburgh, Scottish Academic Press).

PRYOR, J. & TORRANCE, H. (1996) Teacher-pupil interaction in formative assessment: assessing the work or protecting the child?, *The Curriculum Journal*, 7, pp. 205-226.

RAMAPRASAD, A. (1983) On the definition of feedback, *Behavioral Science*, 28, pp. 4-13.

SADLER, D.R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, pp. 119-144.

Säljö, R. (1979) Learning about learning, *Higher Education*, 8, pp. 443-451

SCAA/ACAC (1995) *Consistency in Teacher Assessment: exemplification of standards. English: speaking and listening*, video and booklet (London, SCAA)

SCRE (1995) *Taking a Closer Look at Science* (Edinburgh, Scottish Council for Research in Education)

SOED (1993) National Guidelines, *Environmental Studies 5-14* (Edinburgh, Scottish Office Education Department)

TUNSTALL, P. & Gipps, C. (1996) Teacher feedback to young children in formative assessment: a typology, *British Educational Research Journal*, 22, pp. 389-404.