Developing synergies between formative and summative assessment

David Carless

Associate Professor, Faculty of Education, University of Hong Kong
dcarless@hkucc.hku.hk


Draft – work in progress

Abstract

Formative assessment risks being swamped by the power of summative assessment. In order to flourish, initiatives which support positive links between formative and summative functions of assessment are sorely needed (Broadfoot & Black, 2004; Harlen, 2005). This paper proposes two such strategies: pre-emptive formative assessment (Carless, 2006, 2007a) which involves teacher actions to clarify student understandings before misconceptions have resulted in ineffective learning outcomes and/or loss of marks in summative tests; and the formative use of summative assessment (Black et al., 2003). Case study data from a primary school in Hong Kong are used to exemplify the concepts. Implications are outlined with respect to the development of synergies between formative and summative assessment and the implementation of formative assessment in Chinese contexts.

Introduction

The potential of formative assessment for enhancing student learning is now well-recognised (e.g. Black & Wiliam, 1998) but obstacles in wide-scale implementation remain. A key challenge is the influence of summative assessment looming over formative assessment. Cultures of testing and accountability may crowd out formative assessment or prompt teachers to downplay it. Even for teachers buying into the principles of formative assessment and trying to implement them, they may be constrained by the different, often contradictory demands of the external test system (Black & Wiliam, 2004). If formative assessment is to flourish, initiatives aimed at supporting a positive link between formative and summative are sorely needed (Broadfoot & Black, 2004; Harlen, 2005). No easy task in view of previous unsuccessful attempts to design a system of National Curriculum Assessment that was simultaneously formative, summative and evaluative (Black, 1997).

This paper explores potential synergies between formative and summative functions of assessment at a conceptual level and at the level of practice. The first part covers the relationship between formative and summative functions of assessment; pre-emptive formative assessment and the formative use of summative tests. The second part provides some contextual background to the Hong Kong primary school context and then uses small-scale case study data drawn from classroom observations, face to face and e-mail interviews, relevant M.Ed assignments and analysis of school artefacts, such as test papers and examination results.

The relationship between formative and summative assessment

The extent to which formative (FA) and summative assessment (SA) are part of the same thing or should be kept separate has been discussed in a number of papers (e.g. Biggs, 1998; Harlen & James, 1997; Taras, 2005). Formative and summative assessment are usually distinguished in terms of function and purpose. The former is about aiding learning, the latter has a primary function of grading or measuring. Wiliam & Black (1996) observe that all assessments have the potential to serve a summative function, but only some have the additional capacity of serving formative functions. They further argue for separating the elicitation of evidence from the interpretation of the evidence. This point is reaffirmed by Harlen (2006) distinguishing formative and summative assessment in terms of how the evidence is used.

Knight & Yorke (2003) propose reasonably that the intentions should be separate and whether a task is formative or summative should be transparent to students. Harlen (2006) suggests that information gathered by teachers could be used for both formative and summative purposes. It might however, be considered unhealthy or unfair to use information from a formative task for a summative purpose. One finds support here from Scriven (1967), generally seen as the originator of formative/summative terminology, when he states “formative evaluators should ideally exclude themselves from the role of judge in the summative evaluation” (p. 45). In reality, however, we are consistently making judgements and it may be difficult to erase such observations from our memory and in any case such judgements are part of life and for example, this is probably how most staff appraisals operate.

An argument for separating them is that SA is more powerful and that FA risks being swamped by SA. As Harlen (2005) puts it, if we fuse formative and summative, the latter will dominate. But this can also be an argument for synergy in that formative assessment has a relationship with its powerful partner and rival summative assessment. Biggs (1998), in his critique of the Black and Wiliam (1998) review, argues that there is a strong interaction between FA and SA and a powerful way to enhance learning is to engineer positive backward from SA so as to support feedback from FA. This is similar to the line taken by Taras (2005) who puts it succinctly as follows: “FA is in fact SA plus feedback which is used by the learner” (Taras, 2005).
Kennedy et al., (2006) argue that the bifurcation between FA and SA is no longer useful and that there is a need to rehabilitate SA so that its potentially negative impacts do not detract from the positive effects of formative assessment. Whilst SA and FA should continue to be different in form and function, if SA can deploy some of the principles informing FA then it can exert a positive effect on teaching and learning strategies (Kennedy et al., 2006). Elsewhere with respect to assessment in higher education, I have used the term ‘learning-oriented assessment’ to represent ‘good assessment’ that attempts to transcend formative/summative distinctions (Carless, 2007b; Carless, Joughin & Mok, 2006; Keppell & Carless, 2006).

In summary, formative assessment cannot be viewed in isolation from summative assessment. Praising the virtues of formative assessment and seeing the negative impacts of summative assessment ignores the realities of teachers’ and students’ worlds. An ideal is for formative and summative tasks to be profitably used to inform or support each other. The potential ways forward discussed in this paper are essentially twofold. The first involves the use of formative assessment to support students’ preparation for summative assessment. For example, Brookhart (2001) reports that good students try to use all assessments formatively and often draw on formative and summative data in making adjustments to their work. Secondly, feedback from a summative task has potential to be used to inform current and future learning. This is often not the case because of the finality of a grade (Boud, 1995) and its ego-invoking nature (Butler, 1988). Black et al., (2003) discuss this in terms of the formative use of summative tests and this is explored further in a later sub-section.

Pre-emptive formative assessment

Conceptually, the first potential synergy I term ‘pre-emptive formative assessment’, a strategy which I argue is common in practice but not articulated fully in the existing literature (Carless, 2006). Pre-emptive formative assessment aims to build on traditional practices, such as mock tests but strives to make them into more of a deep learning experience. It also seeks to extend the range of practices falling under the umbrella of formative assessment and so encourage wider implementation (see figure 1 in Appendix 1). Timing and student engagement are seen as key concepts, explicitly seeking to address the problem that much feedback occurs too late for it to be acted upon optimally. Pre-emptive formative assessment is defined as teacher actions, which attempt to clarify student understandings before misconceptions have resulted in ineffective learning outcomes and/or loss of marks in assignments (Carless, 2007a). The bridge between pre-emptive formative assessment and summative assessment is that it has the potential to be an effective strategy to prepare students for high-stakes tests. Whilst risking being no more than test preparation, pre-emptive formative assessment seeks to go one step further by focusing explicitly on student difficulties and engaging students actively.

In terms of how pre-emptive formative assessment might be deployed in the primary school context, test preparation provides a likely starting-point. In terms of the use of summative tests preparation can be a form of FA and also lead to a backwash from SA (cf. Biggs, 1998). Pre-emptive formative assessment may be carried out by engaging students in an activity or task which relates to one which they need to do for a high-stakes assessment. For example, the task may be similar or parallel in terms of developing required knowledge or understanding. Or alternatively, the task may involve students in practising a sub-skill which forms part of a larger assessment task. A specific characteristic of the task is that it seeks to remediate difficulties likely to be experienced by students. It aims to be more than just another practice activity by focusing explicitly on tackling problems commonly experienced by learners. In sum, learners carry out a task or tasks in a challenging domain, known to pose difficulties, which mirrors in some way aspects of what is needed for a graded assessment, such as an assignment or test (Carless, 2007a).

The formative use of summative assessment

The second strategy, familiar from the work carried out by Black et al., (2003), is the formative use of summative tests. This strategy seems particularly relevant in contexts such as Hong Kong, where testing is frequent but more formative approaches are still only at a modest level of implementation. A limitation of the strategy is that evidence is less fine-grained than from formative interactions between teachers and pupils; and that students with learning difficulties may be deterred by tests and performance targets (GTC, 2004).

The strategy can also involve using formative assessment practices to support revision and developing more effective reviewing strategies, these share similarities with the pre-emptive formative assessment notion discussed above. Another strategy relates to how information from tests can be fed back into the teaching and learning process so as to enhance student understandings and performance. The case study teachers (Black et al., 2003) used strategies such as rectifying the learning on test questions done poorly by students; peer marking of test papers; and re-working of examination answers. Carter (1997) discusses a strategy of test analysis, whereby students are returned their test paper with no mark on it but some brief indications (e.g. through ticks, circles and crosses) of student success. Before the final mark is awarded, students are required to analyse, critique and correct their work and explain their errors. The strategy was viewed as successful in shifting responsibility to the learner, they achieved a higher awareness of careless errors and these declined on subsequent tests. Students seemed to demonstrate increased retention of knowledge as evidenced by performance on semester and final exams, and develop better attitudes about test taking and learning.

Hong Kong primary school contextual background

Black and Wiliam (1998) point out that assessment is a social process taking part in a social setting. The role of context is important so some discussion of the Hong Kong setting is in order. In Hong Kong teachers are specialists rather than generalists, usually having 1-2 main subjects and they may also be required to teach 1-2 other subjects. Up until the mid 1990s primary school teaching was mainly a non-graduate profession, with the main teaching qualification being a two or three year teaching certificate. In recent years, more and more graduate teachers have entered primary schools and teachers are now often studying for Master degrees. In terms of pedagogy, teaching is generally traditional, text-book oriented and test-driven (Adamson & Morris, 1998). There is generally an emphasis on
Local primary schools have a long tradition of operating hierarchically with school principals employing mainly top-down decision-making (Morris & Lo, 2000). The recent introduction of Curriculum Leaders represents a new middle management position with roles including, supporting curriculum reform, developing the curriculum across a school and promoting a culture of professional exchange within schools (Lo, 2007). Parents are generally active in their influence on schools and are perceived to demand a lot of homework and testing.

Schooling in Hong Kong has tended to be competitive with norm-referenced examinations playing a determining role in students’ life prospects (Pong & Chow, 2002). In recent years, falling birth rates have increased competition amongst schools with threats of school closure intensifying pressure on principals and teachers. At the same time, there are also initiatives to try to mitigate unwanted side-effects of a stressful culture of testing. In terms of assessment developments, two initiatives are particularly relevant. Firstly, government reform proposals have put forward the notion that more attention should be paid to ‘assessment for learning’. The available evidence, although not extensive, indicates that the implementation of formative approaches to assessment in Hong Kong schools is sparse (Carless, 2005; Cheung, 2001; Morris et al., 2000; Yau, 2004).

Secondly, Basic Competency Assessments (BCAs) have been introduced for the three main subjects of Maths, English and Chinese. Basic Competencies are the essential knowledge and skills or basic standards required by students in relation to the learning targets and objectives set out in the curriculum for each key stage (CDC, 2006). BCA comprises two elements, the Student Assessment, a web-based resource bank, operational since mid-2003, provided to assist teachers in developing and selecting the appropriate assessment tasks for their students; and the Territory-wide system assessment (TSA) administered by the government at the three levels of Primary Three (first conducted in mid-2004), Primary Six (first conducted mid 2005) and Secondary Three (first conducted mid-2006) and comprising paper-and-pen mode with an oral assessment component for the two languages. CDC (2006) claims that the TSA is low stakes in nature. Its stated purposes are to provide feedback to schools about their standards in the three subjects of Chinese Language, English Language and Mathematics, so that schools could draw up plans to increase effectiveness in learning and teaching. The territory-wide data also help the Government to review policies and to provide focused support to schools (CDC, 2006).

The stated functions of BCAs are to help teachers and parents diagnose students’ learning, identify their strengths and areas for improvement so that timely assistance can be provided. In other words it seeks to fulfil both formative and summative assessment functions. The reality however, is a focus on statistics when results are published, and teachers find it difficult to channel the feedback data into classroom implementation (Wan, 2006). This further exemplifies the scenario outlined earlier of formative assessment being drowned by summative assessment.

School case study data

Case background

The school case study data below draws on the experiences of Alison, a committed English teacher enrolled for an M.Ed programme. Alison has 8 years of teaching experience, with 6 of them being in the current school. Her main teaching subject is English, a second language for Hong Kong students and one of the key subjects, along with Chinese and Mathematics. The school is a typical Hong Kong primary school which has been operating since 1998. There are 5 classes in each year level from primary 1 (year 1) to primary 6 (year 6) with a total of 1030 students and 50 teachers in the school. The case under discussion involves the teaching of a primary 1 (year 1) class. There are 8 English lessons per week lasting 35 minutes each.

Innovation and origins

The particular assessment innovation introduced in the school was a mandated period of two “Consolidation Days” after a test. The previous practice had been to go through the answers to the test and then continue with the teaching schedule. The innovation was for teachers to design and implement some follow-up activities for the purpose of review, consolidation and extension. The origins of the innovation were threefold and largely externally generated. The Curriculum Leader had attended a course in which this strategy was recommended; Alison had also attended an in-service course introducing her to the philosophy and practice of Assessment for Learning. The third source of origin was related to the element of the TSA which provided feedback to schools on test results. Part of the rationale for the innovation the case was a desire on the part of the school to use feedback on test results to improve student performance in the TSA.

Test design and processes

Test design and frequency requires some discussion. There were 8 ‘testing events’ (as represented in table 1) in a school year with a set passing mark of 60% (set fairly arbitrarily based on general experience). Although the students treat the testing events seriously, there are no direct consequences for them (except parental or teacher praise or disapproval).

Table 1. Testing events for the school year
Supporting less able students

One of the main potentials of the strategy being adopted was seen as being its potential to support the less able students. As Alison puts it, "for the weaker ones, the data collection is valuable and then it is necessary to identify an area for improvement. It has to be focused, you can't improve everything, you have to choose a feasible area; helping the weaker ones not make so many mistakes is important". In terms of pre-test revision for the less able, Alison notes that it is important to limit the scope of what they are trying to learn e.g. focus on spelling selected vocabulary items or memorizing a particular sentence pattern. She notes some positive outcomes, for example, "their attitude is better and they are more active in participation". In terms of pre-test revision for the less able, Alison notes that it is important to limit the scope of what they are trying to learn e.g. focus on spelling selected vocabulary items or memorizing a particular sentence pattern. She notes some positive outcomes, for example, "their attitude is better and they are more active in participation". My interpretation was that the weakest students seemed to make the most gains from the systematic pre- and post-test interventions. This seems to provide some support from a Chinese context of one of the findings summarized in Black and Wiliam (1998) that formative assessment was found to be particularly successful for raising standards amongst the lower ability/performing students. Alison also pointed out that another factor was that the higher ability students were more aware of their learning needs.
Allison mentions some strategies which support the less able students. She notes that oral feedback is more useful than written feedback through marking, as she puts it, “speaking individually to the weaker ones is a good strategy. The individual approach can help them build confidence”. The quiz before the test can also help them to focus their learning and if they are reasonably successful can play a role in building up their confidence. She also refers to a strategy of teaching them ‘learning skills and revisions strategies’ as follows:

I remember that one of the strategies that I used about their "corrections" was that I asked them to refer back to their books and notes for the answer instead of just going through them like what my colleagues usually do. I think it is a good chance to develop their learning skills. They have to actively "do something" for the answers not just passively "listen" to their teachers. After that, there were two times of writing practices where both Bobby and Heidi [two of the less able students] did try to refer to their story books and notes to find the words or sentence they like to use in their own writing. In the past, they would just ask their teacher directly and "wait" for the answer.

Overall impact of innovation

Overall, Allison felt that the innovation was successful, although it did lead to an increase in teacher workload and it was not particularly welcomed by colleagues. She saw the benefits as being fourfold. Firstly, she identified better reinforcement of learning by students as they were directed to plan their revision more wisely. Secondly, as she puts it, “I can see that at least for the weak ones, they feel more confident about learning”. A third positive aspect was the systematic nature of the analysis of student performance: “It is systematic, other teachers may go just on impression, but impressions are not always accurate, in particular for all students so it may not cater for learner differences”. Fourthly, Allison also viewed the process as a form of professional development and increases awareness about the relationship between curriculum, pedagogy and assessment and how to make good use of test data.

Despite Allison’s positive feeling about the innovation, she reports that the other teachers were somewhat passive and unenthusiastic about the consolidation period (used in Chinese, English and Maths). She attributes this partly to their view that a test is the end of learning and then they go on to another topic. As she puts it, “in their view all testing is summative, they lack the formative concept”. She notes that “Teachers have to be aware of using the results and their implications, using the data and not just reporting it”. In other words, she perceives her colleagues as seeing assessment as something to measure and report, rather than to use for the purpose of enhancing learning. Such a finding is consistent with research into Hong Kong teachers’ assessment practices (e.g. Morris et al., 2000).

There are also negative implications for teacher workload which is already heavy:

Analyzing and interpreting the examination data systematically is a very time consuming process. It took the teachers around an hour just for recording the data for a single class. The staff meeting followed for strategic planning of follow-up actions demanded even more time and effort of the participants. The heavy workload of teachers may hinder them from adopting, or persistently adopting the approach because this is not the only task teachers have to do in their daily teaching role. Even though there is strong evidence that students do improve using the approach and recognition have been gained from their colleagues for their professionalism, it is still difficult for a teacher to persist it when there is pressure of being overload. As one teacher commented, “It is really an effort-demanding process and I might easily be tempted to give up if I do not have enough support from you during those hard times”.

In view of these challenges, Allison perceives that the role of management is particularly important in terms of supporting colleagues. She viewed support from middle management as being crucial e.g. control of workload distribution and some recognition or rewards. In sum, the innovation seemed to have positive potential for pupils in terms of reinforcing their learning and in increasing their confidence. For the teacher, it increased workload, but represented a systematic intervention which contributed to her professional development.

Conclusions

The reconciliation of formative and summative assessment is not a new challenge, but it is a crucial one if formative assessment is to meet its potential. Formative assessment is difficult for teachers to implement. This is the case in all contexts but it may be particularly difficult in a context, such as Hong Kong where summative assessment has tended to swamp approaches of a more formative orientation. In such a situation, if formative and summative are not integrated at all, the formative aspects are likely to be neglected. In Confucian heritage cultures to date, teachers do not seem to have really come to grips with the learning potential of assessment. There is clearly further work that needs doing both in terms of professional development for teachers and research with teachers.

This paper has presented one diligent and well-informed teacher attempting more formative approaches to assessment. In a test-dominated context, the strategy of attempting synergies between formative and summative assessment seems a promising one. Or to put it another way, if schools are going to spend a lot of time on testing, it is essential for students to derive some solid learning from the test process, before, during and after. Two particular strategies have been discussed. Pre-emptive formative assessment represents an attempt to tinker with exam practice and make it a learning process that promotes timely input and student engagement. The formative use of summative tests seeks to use data from test results in a productive way to enhance student learning.

References


*This document was added to the Education-Line database on 15 December 2006*