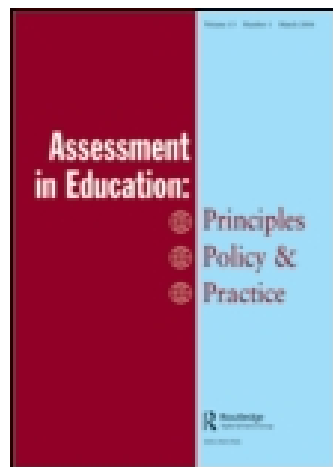


This article was downloaded by: [University of Bath]

On: 15 April 2015, At: 03:23

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment in Education: Principles, Policy & Practice

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caie20>

Assessment and Classroom Learning: a role for summative assessment?

John Biggs^a

^a University of New South Wales, Professional Development Centre, PO Box 114, Killcare, New South Wales 2257, Australia

Published online: 28 Jul 2006.

To cite this article: John Biggs (1998) Assessment and Classroom Learning: a role for summative assessment?, *Assessment in Education: Principles, Policy & Practice*, 5:1, 103-110, DOI: [10.1080/0969595980050106](https://doi.org/10.1080/0969595980050106)

To link to this article: <http://dx.doi.org/10.1080/0969595980050106>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Assessment and Classroom Learning: a role for summative assessment?

JOHN BIGGS

*Professional Development Centre, University of New South Wales, PO Box 114,
Killcare, New South Wales 2257, Australia*

Introduction

The very comprehensive and carefully researched review by Black & Wiliam addresses the issue dear to every educator's heart: 'Significant learning gains lie within our grasp.' (p. 61). The authors put on display a wealth of research that does indeed suggest that this statement is true; they see the clue as the proper utilisation of *feedback*. This they interpret liberally to refer to 'any information that is provided to the performer of any action about that performance'. (p. 53); the word 'assessment' in the title refers only to formative assessment (FA).

However, it is well into the article before this position becomes clear. In the introduction, two main purposes are outlined: to look at the contribution assessment makes towards better learning, and to attempt a synthesis between theoretical and practical issues. The starting point is effectively 1988, prior to which are the similar reviews by Natriello (1987) and Crooks (1988). They state that Crooks used the term 'classroom evaluation' with the same meaning as we propose for 'formative assessment' (p. 8), which is not quite true as Crooks also had much to say about summative assessment (SA). This leads to the main weakness of their review, as discussed below.

The Structure and Content of the Review

An evaluation of Black & Wiliam's review requires a quick run through in order to get a feel for their argument. They start in an interestingly different way, with a selection of seven studies, which set the agenda, and provide some key foci. Following sections then address: Teachers' practices, students and formative assessment, strategies and tactics for teachers, systems, feedback, and prospects for the theory and practice of formative assessment.

The section on teachers' practices mostly endorses what we already knew from Crooks' review: teachers tend to use practices that encourage low cognitive level activities such as recall of isolated items of knowledge, much testing is unreflective and not aligned to the curriculum, and teachers see assessment as norm-referenced so that the effect of feedback on low performers is to tell them that they lack ability. National or local requirements for certification and accountability greatly affect practise, requiring teachers to rethink their own role and their conceptions of

teaching and assessment, which they do not seem to be very good at doing. This last point highlights an important focus for pre- and in-service teacher education and staff development, which is picked up in the tertiary literature at least (e.g. Ramsden, 1992; Marton *et al.*, 1997).

The effectiveness of FA depends on whether students actually perceive the gap between where they currently are and where they should be; and then if they do, what they are willing to do about closing it. This brings in a great deal of work on learning and performance goals, and the attributions usually associated with them. There is a constructivist subtext that students are intrinsically and necessarily involved in receiving and acting upon feedback. The research reviewed supports the authors' position that making feedback effective can enhance student learning, and that handing over responsibility for assessment to the students is particularly effective. Yet in most classrooms, even in tertiary classrooms, self-assessment is rare.

The next section addresses the pedagogy for making feedback more effective. The review focuses on: choice of task, classroom discourse, use of questions by both teacher and student, level and frequency of tests, quality of feedback, and the need for a framework for formulating strategy, the last involving 'collaboration between psychometricians, cognitive scientists, and subject experts' (p. 39) (but hopefully not in that order of priority—see below).

The section on systems is probably the weakest, much time being spent discussing mastery learning, and assessment driven models. The conclusion is drawn that 'the only clear messages emerging from the mastery learning literature are that mastery learning appears to be effective in raising students' scores on teacher-produced tests ... in teacher-paced programmes than in self-paced programmes, and ... for younger students' (p. 42). This ignores the issue of learning quality, which unfortunately tends to happen if we talk only about 'effectiveness' in a quantitative kind of way. For there is another message from that literature; mastery learning is 'effective' with students who are prone to a surface approach to learning, and in encouraging low level outcomes, but singularly ineffective with deep-oriented students (Lai & Biggs, 1994).

The remainder of this section deals with curriculum based assessment (CBA), its quantitative derivative curriculum based measurement (CBM), and with portfolio assessment, with its high order overtones of reflective thinking, but little hard research data. I found it difficult to follow a continuing thread here, one reason being I think that the dimensions were crossed: a point I come back to in considering Table 1 (see p. 60), an adaptation of Shinn & Hubbard's (1992) CBA model (see below). Accordingly, I was left with little more than the conclusion that I already held strongly: classroom assessment that *is not* 'curriculum based' is bad assessment, whether FA or SA.

In CBA, 'the main purpose for assessment is the formative purpose' (p. 44), but that by no means precludes SA. In some versions of CBA, such as Hong Kong's Target Oriented Curriculum (TOC) (Education Department, 1994), students' performances on the target assessment tasks are used to plot their course of attainment growth on an individual basis; information pertaining to how far and in what way an individual student falls short of a given criterion-referenced target is

used formatively in teaching. As for SA, the targets are arranged in eight bands of performance that accumulate over four 'key stages', so that at any given time, individual students find themselves placed in a band of performance, within a key stage, so that the same set of assessments also serve the summative function. This longitudinal or growth model of assessment thus seems to achieve both formative and summative functions.

The problem with this is not to do with the structure of the assessment, but with Hong Kong's strong tradition of rigorous norm-referenced summative assessment. Throughout primary school, but grimly so in the Primary 5 and 6 years, students are trained in test-taking skills so that they perform optimally in the Secondary Schools Placement Allocation, the results of which in effect determine the quality of their secondary schooling, and indeed post-secondary and life options. In such a system, competition is fierce, teachers and parents uniting to push students as far as they can. Teaching is carried out in large classes, and is narrowly focused on maximising external assessment results, while students see memorisation as a core learning strategy (Biggs, 1996c). In Western educational wisdom, this constitutes an impoverished teaching/learning environment, yet Hong Kong students, and students from Confucian heritage cultures in general, not only outperform Western students but do so using deep, meaning-oriented, learning strategies, thereby creating 'the paradox of the Chinese learner' (Watkins & Biggs, 1996). The so-called paradox is embedded deep in the cultural supports for learning, which are often dimly perceived by Western eyes; certainly generalisations to Western culture (e.g. 'large classes are OK after all!') are quite out of place (Biggs, in press).

The participants in the Hong Kong educational system thus have very strong existing beliefs in the moral and functional value of norm-referenced summative assessments, from the weekly test to the SSPA (Secondary School Placement Allocation [Hong Kong 11 +]). The existing system of secondary allocation thus rides uneasily alongside the CBA-derived TOC itself. Something will have to give.

These problems are not peculiar to Hong Kong. That system is only a good illustration of two universal issues. The first is that sensible educational models of assessment make effective use of both FA and SA, and that is dealt with below. The second is that the psychometric model of assessment confuses the educational issues, and that is dealt with in the section after that.

The structure of Black & Wiliams' review, then, is loose and sometimes repetitive, while there are some arguable content issues, as elaborated below. It was an interesting try, but a top-down treatment would have been more effective, and had the definitions of FA, SA, and feedback had been addressed at the beginning rather than at the end of the paper, at least one reader would have had a clearer sense of direction.

The Relationship between SA and FA

My main substantive problem is with the decision to exclude SA from a review of the effects of assessment on classroom learning. This is not just a reflection of my own fancies, but is intrinsic to the central issue of obtaining significant learning gains

through assessment practice. From the outset, the effects of FA are seen as positive: 'we have not come across any report of negative effects following an enhancement of formative practice' (p. 17). Untangling the conditions under which this optimally takes place directs this article, and that is certainly useful. However, it leaves in abeyance what some would see as an even bigger question, the effects of SA on learning.

These effects, referred to as 'backwash' (Biggs, 1996b), are usually seen as entirely negative, and interestingly, as stronger than the positive effects of FA ('feedback'). This suggests clearly that significant gains are to be found as much in mitigating or reversing backwash as by enhancing feedback. Several studies reviewed here—Butler (1988), and many studies by Schunk and Ames and their respective teams into process/product goals, and learning and performance goals—bear this out. In Butler's study, for instance, three groups of students were given comments only (straight FA), comments and final grade (FA + SA), and grade only (SA). The FA group's performance on two kinds of task increased, the FA + SA group's performance fell significantly on both tasks, the SA group's fell then rose in one task. These results are interpreted plausibly in terms of cognitive evaluation: the grade helped define the comments as ego-related and not task-related.

But whatever the explanation, there is a powerful interaction between FA and SA that could usefully be incorporated in an overall synthesis, so that both backwash (from SA) and feedback (from FA) are conceptualised within the same framework. Then indeed one might have a powerful enhancement to learning, using such a synthesis to engineer backwash from SA so that the effects were *positive*, the backwash from SA *supporting* the feedback from FA, instead of as in Butler's study and other studies reviewed here, nullifying it.

Why SA has been ruled out from this review, we only learn in the last section on synthesis. Feedback concerning the gap between what-is and what-should-be is regarded as 'formative' only when 'comparison of actual and reference levels yields information that is then used to alter the gap' (p. 53). But if the information cannot lead to appropriate action—it becomes a summative grade, for instance—then it is not formative (after Sadler, 1989). And that is the basis for the distinction between FA and SA; they are seen in effect as mutually exclusive.

Whether they are mutually exclusive or not depends on how inclusive one's model of assessment is. Let us look at the backwash (SA) and feedback (FA) effects on learning, and let us characterise the latter in terms of a student's *approach* to learning, a concept that has considerable currency, particularly but not only in the tertiary literature (e.g. Biggs, 1993; Marton *et al.*, 1997). The backwash from SA is generally agreed to be negative, generating ego-related and other non-task priorities that in turn create a surface approach, where the learner uses lower cognitive level activities than those actually required, so that performance is low level and fragmentary. The feedback from FA on the other hand facilitates learning, providing the information needed for a deep approach, which involves deploying cognitive activities that are appropriate to the level required by the task, as depicted in Figure 1.

The feedback from FA derives from P2, the backwash from SA from P3. The

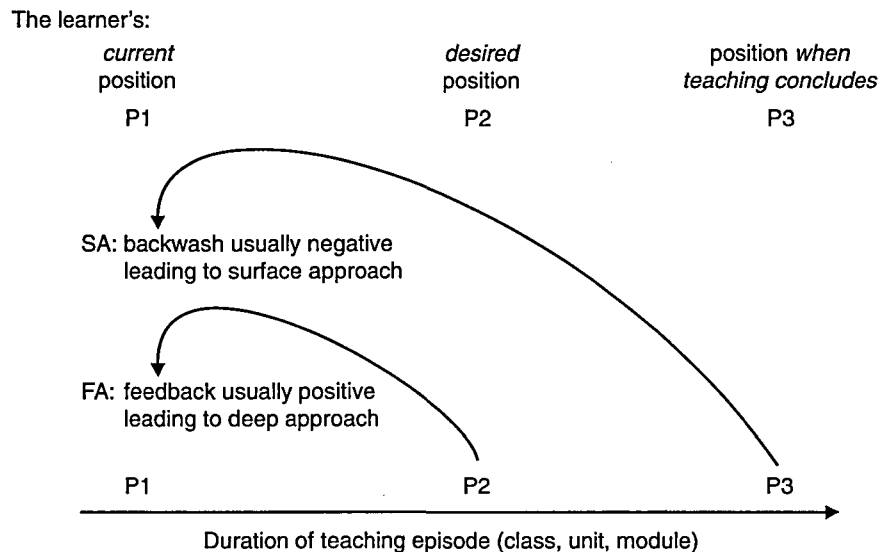


FIG. 1. Formative and summative assessment during the course of learning.

latter comes about because the students perceive the assessment task demands, and as far as they are concerned 'assessment always defines the actual curriculum' (Ramsden, 1992, p. 187). Such strategies as cue-seeking (Miller & Parlett, 1974) are their ways of obtaining FA from teachers, and of course when teachers are teaching to the test, they are quite upfront about their role in providing feedback ... or is it backwash from SA? Distinctions at this stage are becoming fairly semantic: 'When the chef tastes, it's formative assessment; when the customer tastes, it's summative' (source regrettably forgotten).

Figure 1 suggests that students' approaches to learning mediate the effects of both feedback (FA) and backwash (SA) on performance (see also the presage-process-product model in Biggs, 1993), so that differences between the two are not matters of principle so much as of timing. If we take an information-rich mode of SA, such as the assessment portfolio graded qualitatively, these differences converge still further. The portfolio receives little sympathy in studies reviewed by Black & Wiliam, but when it is used summatively and designed appropriately, it is very good at setting in motion metacognitive and reflective learning processes that generate much feedback (Biggs, 1996a,c, see ch. 9); students themselves, with or without teacher help, can focus on the gap between their actual position (P1) and the desired position (P2), until the calendar dictates P3, when teaching must stop.

In this situation, which is characteristic where reflective learning takes place, the backlash from the summative assessment tool can be very positive. A condition is that the assessment is deeply criterion-referenced, incorporating the intended curriculum, which should be clearly salient in the perceived assessment demands. When

that happens, you get aligned instruction, where teaching to the test is exactly what you want because it is teaching the intended curriculum (Biggs, 1996a). In other words, when the *summative* assessment is defining the parameters for the *formative* assessment, it does not seem helpful to confine a review of the one with the attempted exclusion of the other. I say 'attempted', because the interaction between the two means that it is very difficult in the event to deal with FA to the exclusion of SA, and as we have seen here, SA does enter into many of the studies under disguise: for example as a 'performance goal', which students interpret as they would a summative test.

It does not even remain true that at P3 'the control loop cannot be closed' (quoted as Sadler's (1989) deeply coded way of distinguishing SA from FA). If the grade awarded is based on a qualitatively derived hierarchy, such as one using SOLO or any other suitable taxonomy based on the evolution of learning (Biggs, 1992, 1998b), then that grade becomes highly informative, and although one may not be able to do much about it with regard to the present teaching session, it can certainly provide very valuable feedback to the student for the next.

I may be seen here as committing the reviewer's cardinal sin: reviewing the article that I would have liked to have written, rather than the one on the agenda, which is what somebody else actually has written. I hope my comments can be seen as deriving from a similar starting point as Black & Wiliam, but instead of seeing FA and SA up close as two different trees, I would zoom to a wider angle conceptually. Then, in the broad picture of the whole teaching context—incorporating curriculum, teaching itself (an excellent feature of their review), and summative assessment—instead of two tree-trunks, the backside of an elephant appears.

Other Issues

The Problem of Psychometrics

The wide angle shot of the instructional elephant addresses another problem: seeing psychometricians, and cognitive scientists and subject experts, as the ones to address the task of developing 'cognitively diagnostic assessments'. Classical psychometrics had been directed towards problems of selection, whereas the present need is to use assessments to guide learning (Nichols, 1994). True, but I do not think we should look for help to psychometricians, who through no fault of their own have nonetheless done enough damage to educational assessment already.

Let me explain this harsh statement. Psychometricians are concerned with the quantification of individual differences, and they have constructed a massive and sophisticated technology to that end. Educators are concerned with something entirely different, but the language, actions, and procedures of many teachers and more administrators, and the conceptions of assessment they hold, derive from psychometrics. Those psychometric conceptions actually fight the conceptions needed to understand and implement criterion-referencing and other attributes of sensible instructional design (Sheppard, 1991; Gipps, 1994; Taylor, 1994; Biggs, 1996a,b). No, a paradigm shift is surely necessary, but let us look elsewhere than to

psychometricians to provide it, because their assumptions are simply inappropriate to the educational purpose.

Teachers and educators are primarily interested in bringing about *change* in students' performances. How can you address change using concepts and a technology based on the stability of traits, and their normal distribution in the population? And how can qualitative changes in cognitive structure be reduced to a point along a continuum, even if you do call that point a 'mark', or the continuum a 'percentage scale'? Black & Wiliam draw attention to the need for teachers' conceptions to change, but I am disappointed they did not follow up with a discussion of the dimensions most in need of change. Measurement model *vs* standards model (Taylor, 1994), and quantitative *vs* qualitative (Cole, 1990), are two critical dimensions of teacher thinking that need addressing head on. Much of this argument is in Gipps (1994), who is mentioned in the review, but her generic points do not seem to have been taken on board.

What are the Key Dimensions?

Finally, I would take issue with Shinn & Hubbard's (1992) list of polarities, which are reproduced in Table 1 (see p. 60, this issue) as the way to determine the essential elements of any strategy to improve learning through FA. This consolidates what I see as the main problems of the review itself—dimensional ambiguity. Summative *vs* formative, spreading out individuals *vs* socially meaningful outcomes, groups *vs* individuals, unobservable constructs *vs* target behaviours or skills, stability *vs* variability in performance, and so on, are laid out as if they address a single underlying generic polarity. To call this confusion of dimensions a 'paradigm shift' is not helpful.

In sum, then, the authors are to be congratulated for a thorough review and compilation of studies that do address what is after all their central purpose; to examine the conditions under which feedback from FA may be maximally effective. My main disagreement is that I would rather see FA in a broader context, embracing a multidimensional view of the instructional process.

References

- BIGGS, J.B. (1992) A qualitative approach to grading students, *HERDSA News*, 14(3), pp. 3–6.
- BIGGS, J.B. (1993) What do inventories of students' learning processes really measure? A theoretical review and clarification, *British Journal of Educational Psychology*, 63, pp. 1–17.
- BIGGS, J.B. (1996a) Enhancing teaching through constructive alignment, *Higher Education*, 32, pp. 347–364.
- BIGGS, J.B. (1996b) Assessing learning quality: reconciling institutional, staff, and educational demands, *Assessment and Evaluation in Higher Education*, 21, pp. 3–15.
- BIGGS, J.B. (Ed.) (1996c) *Testing: to educate or to select? Education in Hong Kong at the crossroads* (Hong Kong, Hong Kong Educational Publishing Co.).
- BIGGS, J.B. (1998a) Learning from the Confucian heritage: so size doesn't matter, *International Journal of Educational Research*, in press.
- BIGGS, J.B. (1998b) *What the Student Does: teaching for quality learning in higher education* (Buckingham, Open University Press), in press.

- BUTLER, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on student interest and performance, *British Journal of Educational Psychology*, 58, pp. 1-14.
- COLE, N.S. (1990) Conceptions of educational achievement, *Educational Researcher*, 19(3), pp. 2-7.
- CROOKS, T.J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, 58, pp. 438-481.
- EDUCATION DEPARTMENT (1994) *General Introduction to Target Oriented Curriculum* (Hong Kong, Government Printer).
- GIPPS, C. (1994) *Beyond testing: towards a theory of educational assessment* (London, Falmer).
- HARRIS, D. & BELL, C. (1986) *Evaluating and assessing for learning* (London, Kogan Page).
- LAI, P. & BIGGS, J.R. (1994) Who benefits from mastery learning?, *Contemporary Educational Psychology*, 19, pp. 13-23.
- MARTON, F., HOUNSELL, D. & ENTWISTLE, N. (Eds) (1997) *The experience of learning* (Edinburgh, Scottish Academic Press).
- MILLER, C. & PARLETT, M. (1974) *Up to the Mark* (Guildford, Society for Research into Higher Education).
- NATRIELLO, G. (1987) The impact of evaluation processes on students, *Educational Psychologist*, 22, pp. 155-175.
- NICKOLS, P.D. (1994) A framework for developing cognitively diagnostic assessments, *Review of Educational Research*, 64, pp. 575-603.
- RAMSDEN, P. (1992) *Learning to Teach in Higher Education* (London, Routledge).
- SADLER, R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, pp. 119-144.
- SHINN, M.R. & HUBBARD, D.D. (1992) Curriculum-based assessment and problem-solving assessment: basic procedures and outcomes, *Focus on Exceptional Children*, 24(5), pp. 1-20.
- SHEPARD, L.A. (1991) Psychometricians' beliefs about learning, *Educational Researcher*, 20(6), pp. 2-16.
- TAYLOR, C. (1994) Assessment for measurement or standards: the peril and promise of large scale assessment reform, *American Educational Research Journal*, 31, pp. 231-262.
- WATKINS, D. & BIGGS, J. (Eds) (1996) *The Chinese Learner: cultural, psychological, and contextual influences* (Hong Kong, Centre for Comparative Research in Education/Camberwell, Vic, Australian Council for Educational Research).