WCES-2010

Assessment for learning: assessing the theory and evidence

Maddalena Taras* *

“University of Sunderland, Department of Education, Chester Road, Sunderland SR1 3SD, UK

Received October 26, 2009; revised December 2, 2009; accepted January 13, 2010

Abstract

Assessment for Learning (AfL) and its various permutations are used in educational institutions across the world and extensively supported by empirical research (Stobart 2008). However, Stobart observes “given that AfL has largely been driven by academics, it is surprising to find that the theoretical underpinnings have been very much in the background” (Stobart 2008 p149). This paper addresses this deficit in theory and its impact on the AfL framework. It examines the theoretical evidence which can support AfL. A much more solid theoretical basis exists than is perceived by the literature, and a clearer delineation of this will provide increased credibility for AfL within the education community.

© 2010 Elsevier Ltd. Open access under CC BY-NC-ND license.

Keywords: Assessment; theory; formative; summative; assessment for learning.

1. Introduction

Assessment for Learning (AfL) and various permutations of this can be found in use in educational institutions across the world: it began as a distinct movement in the UK which is based on principles to support learners through assessment. AfL has disseminated four interventions: questioning, feedback through marking, peer- and self-assessment, and formative use of summative tests. It is founded on empirical research based on promoting learning from assessment. It is associated with the Assessment Reform Group and the work of Black and Wiliam (Stobart 2008 p147).

It is difficult to conceive of any major educational movement which has inspired academics and teachers worldwide which is not firmly grounded in theory: but this is the reality of AfL. Initially, this paper will examine the claims for theory and how comprehensive or not this theory claims to be, how these claims are then changed and attenuated, and how this theory is perceived by those proposing it and those external to it. A theory which is compatible with the aims of AfL will be presented.

* Maddalena Taras. Tel.: +191-515 2910
E-mail address: maddalena.taras@sunderland.ac.uk

© 2010 Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2010.03.457
2. Assessment for Learning: an a-theoretical practice?

Claims for theory are made by the central instigators of AfL and their academic position is considered sufficient grounding from which to make the claim:

“The authors of this book worked in a university and can lay claim to expertise in the areas of evidence and theory” (Black et al 2003 p1)

In AfL, any serious issues such as reliability and validity, are not considered because AfL does not use assessments for accountability or certification: it does however have an important educational purpose – that of promoting students’ learning and teaching (Black et al 2003 p2).

“…. it is usually informal, embedded in all aspects of teaching and learning, and conducted by different teachers as part of their own diverse and individual teaching styles” (Black et al 2003 p2)

Black et al 2003 is important because in the UK this book is frequently used as a basic text for Post Graduate Certificate in Education programmes which prepare graduates for teaching. The study of AfL is central to these programmes. Black et al 2003 claim to situate AfL within theories of learning, motivation and the role of feedback, but can find no “comprehensive theory” to support their innovation:

“The principles to do with teacher change, student change, student change and feedback would clearly have to be borne in mind in any innovative development. While these pointed to several theoretical ideas that would be relevant, notably those concerned with theories of learning, theories of motivation and Sadler’s analysis of the role of feedback, there is no comprehensive theory that could form a basis for action” (Black et al 2003 p15)

In a book which claims to promote formative assessment and by academics who also claim to have expertise in areas of evidence and theory, there is very little situating of formative assessment within a theory of assessment. Theories of learning and motivation are, of course, relevant, but surely formative assessment should be primarily situated within a theory of formative assessment and a theory of assessment in general: this would require explicit definitions of all the principal elements and argue cogently how these relate and interact.

The above quote is even more puzzling since Sadler is mentioned only as proffering the role of feedback when in reality, his paper (1989) from which they take Sadler’s ‘analysis’ is in fact one of the most cited papers for providing a coherent theoretical framework of formative assessment and how this links to both learning and feedback.

By looking at early work of Black and Wiliam (Wiliam & Black, 1996; Wiliam, 2000; Black et al., 2003) it is possible to understand how many of the difficulties manifest in AfL theory came about. Taras (2007) observes:

“….summative assessment tests ‘are designed to judge the extent of students’ learning of the material ... for the purpose of grading, certification etc.’ and formative assessment is ‘so useful in helping them improve what they wish to do’ (Wiliam, 2000 p18). The focus in both definitions by Wiliam is on the process; in the verb ‘to judge’ for summative assessment, and ‘another type of evaluation’ for formative assessment, yet, as shall become evident, when discussing definitions, it is repeatedly made clear that the distinction is ever only of function or purpose (the two words are used interchangeably): ‘However, if assessments were designed only for summative purposes, then formative information could not be obtained, since the summative assessments occur at the end of a phase of learning and make no attempt at throwing light on the educational history of the pupil’ (Wiliam and Black 1996 p544)” Taras 2007 p364-365

This is the first major problem: that the process of assessment, be it summative or formative, is constantly relegated to oblivion by insistence on the ‘functions’ – consideration of which must take precedence. But what are functions or purposes if not uses that are often social and political, may be multiple, difficult to categorise and even more difficult to contain and respect (Taras 2005)? Assessors may be hypocritical and say that they will be assessing for example, formatively and then use it summatively (Wiliam and Black 1996, Black and Wiliam 1998). What then would be its function? Basing summative and formative assessments on functions is the most disorientating aspect of discussions in the AfL literature and contributes nothing to the understanding of assessment processes.
3. The assessment process

Wiliam notes that ‘Assessment is then considered as a cycle of three phases (eliciting evidence, interpreting evidence, taking action)’ (Wiliam and Black 1996 p537, Wiliam 2000 p1). From this we can summarise that assessment is one single process: the three phases are part of a cycle and thus link up to each other. Taras (2007 p365) notes that it is never quite clear who is involved in either the assessment cycle or uses the information, producing ambiguity as to the assessors and users. The result of this process (including the ‘evidence’ which is presumably the feedback as to whether the work is appropriate or not) becomes summative or formative depending on its use because of:

‘…the need to integrate (or at least align) the routines of informal classroom assessment with more formal assessment practices. It has become conventional to describe these two kinds of assessment as formative and summative assessment respectively, but it is important to note in this context that the terms ‘formative’ and ‘summative’ do not describe assessments—the same assessment might be used both formatively and summatively—but rather are descriptions of the use to which information arising from the assessment is put. (Wiliam 2000 p1, which reiterate Wiliam and Black 1996 p538)

Except that here there is further ambiguity between summative and formative not just with the use of the assessment but also with the context, that is between the informal classroom assessment and formal assessment: as Taras (2007 p61) notes, between process and product. Furthermore,

“Here there is ambiguity between the assessment, i.e. student work, and the assessment, i.e. product of tutor ‘evaluation’; and the use to which the latter is put, i.e. formative or summative assessment. This is in addition to a new differentiation between formative and summative assessment to denote informal classroom assessment, as opposed to formal assessment processes respectively.” (Taras 2007 p365)

If we take the ‘process’ at face value, which we must remember can start or stop at any of these stages (Wiliam and Black 1996 p540), i.e. “eliciting evidence, interpreting evidence, taking action” (Wiliam and Black 1996 p537, how can this cycle fit into a theory of assessment, including process, product and function and be made explicit and clear?

Strange as it may seem, this process is not related in any way to Sadler (1989) and Sadler is used for one definition of formative assessment – that of product assessment, without in any way relating it to the process. One obvious question comes to mind: can ‘evidence’, whether elicited or interpreted be equated with feedback? Logically, I would have assumed that it must, and yet this discussion of ‘process’ is divorced from and kept separate from issues of feedback.

4. Definitions of formative assessment as a process

The second major problem with AfL theory is that definitions of formative assessment are multiple and uncoordinated (Taras 2007 p365). Dual definitions of formative assessment appear across the literature: one is based on Sadler (1989) and has formative assessment focusing on product assessment (Wiliam 2000 p15, Black 2003c p2, Black et al 2003 p15, p121). The other is a classroom learning and teaching pedagogy process (Black 2003a, b, c; Wiliam 1994, 2000a, b, 2000, Wiliam and Black 1998 p8, Black et al 2003 p2).

Even within Black et al 2003 there are two different processes of assessment offered which do not integrate or support each other.

“Such assessment becomes formative assessment when the evidence is used to adapt the teaching work to meet learning needs.” (Black et al 2003 p2)
They continue that formative assessment can cover different methods and different ways of using the feedback and that this can occur many times within one lesson:

“It has to be within the control of the individual teacher and, for this reason, change in formative assessment practice is an integral and intimate part of a teacher’s daily work.” (ibid p2)

From this it follows that formative assessment is part of teaching methodology and has more to do with teachers than learners, thus seeming to contradict the initial definitions. The second process of formative assessment in the book, which seems to be a dramatic conceptual leap, does not seem to be in accordance with the initial definition. When discussing the central concept of feedback (ibid, p13-15) as expressed by Sadler 1989, they note that his comment on feedback which state that it requires information on some attribute, which it compares to the desired level in order to identify a gap; finally the information is used in order to alter the gap.
“With small changes of terminology, the above four steps could be a description of formative assessment.” (ibid p15)

This is Sadler’s definition of formative assessment. This differs from the initial definitions by Black et al and by Black and Wiliam in that firstly, it has to do with learners in that it assesses their work and not teachers’ classroom processes, and secondly, it is about the process of assessment (i.e. four steps) which relates to product and not about teacher methodology in the classroom. Thus this formative assessment does not correspond to their definitions cited above.

What does each of their definitions represent? The product definition taken from Sadler 1989 is straightforward: criteria are used to judge the quality of the work in question. The other of classroom process is much more complex and convoluted and is much more to do with who has decision-making power over whom and who decides what is right and wrong: in a social constructivist model of learning, negotiation of meaning, understanding and action would seem appropriate. This does not seem to be manifested in Black and Wiliam’s definition of formative assessment as classroom interaction despite later claims to links with theories of learning, for example Stobart specifically claims that “the learning theory approach which underpins current AfL positions, including my own, is probably best described as ‘social constructivism’. This seeks to hold in balance learning as a cultural activity and as individual meaning-making” (Stobart 2008 p151).

In addition, they seem to change the definition of formative assessment by equating it to and confusing it with feedback (Black et al 2003 p14-15). Sadler’s article effectively describes and discusses formative assessment, of which formative feedback is an integral part and which takes place when the assessment has been completed: the assessment can be carried out by the tutor or the learner, but the feedback must be used by the learner. It follows, that, to understand the feedback, learners must also understand the difference between the initial assessment and what is required: therefore learners also need to be part of the assessment process. This initial assessment is summative assessment. By not naming the initial step of summative assessment, the whole process of formative assessment seems to have becomes confused by Black et al.

Near the end of the book Black et al note that their work has clarified the concept of formative assessment (ibid p122); this again seems to return to the original definitions cited at the beginning of the section as they have done intermittently throughout the book.

In a book which presents both a theoretical and practical framework of formative assessment, it is surprising that firstly, two separate and disparate processes are presented, secondly, that these are not acknowledged as separate processes or explained in relation to each other, and thirdly, that the interrelationship between the theory and the practice is not discussed.

5. Confusion between Functions and Process

5.1. Formative and summative assessment tensions

Stobart (2008) demonstrates how assessment has been used over the centuries to perpetrate social injustice. Exams for selection are an example of a negative and destructive aspect of assessment which attacks and devalues personal worth and ruins careers: these exams have been categorised as being “summative assessment” in the AfL framework because they perform a “summative” function, and it has been contrasted to formative assessment which is purported to bear none of the negative stigma (Black 2003b, c; Black et al 2003, Wiliam 2000b; Wiliam and Black 1998).

However important these considerations may be in terms of function, this does not provide justification for an artificial division between the processes. This is another example of how the lack of clarity in one area, such as the conflict in ideas of assessment as a function and assessment as a process, can lead to wider confusion and an unnecessary and undesirable conflict between other definitions, in this case formative and summative assessment. If assessment is a single process with more than one function, then the process cannot logically be divided and the process is the same regardless of the function.

If the difference in the kinds of assessment is in terms of process, either by being distinct and opposing processes, or by having one as an extension of the other, then in the first case the division is already logically there and cannot be imposed and in the second case cannot be affected as one kind of assessment by definition would entail the other, regardless of the functions to which they are put. This difficulty in defining and dividing, and the confusion caused by function or purpose driven distinctions is further evidenced in practice. Black et al themselves admit that
“some have argued that formative and summative assessments are so different in purpose that they have to be kept apart. However, our teachers found it unrealistic to practise such separation and so sought to achieve a more positive relationship between the two, at least for those tests where the teacher has control over when and how they are used.” (Black et al 2003 p31)

5.2. Teacher rejection of separation of summative and formative assessment

More recently, Black et al (2003) readjust their ideas on using formative and summative assessment to support each other in the classroom context, essentially because firstly, teachers refuse to separate the two (Black et al 2003 p31) and secondly, because of evidence collected from teachers, who found the usefulness of using summative assessment for formative purposes (Black et al 2003 p55-6).

“Although most of our teachers were able to remove the conflict between formative and summative assessment for some parts of their schemes of work, frequent summative testing dulled the message about the means to improve, replacing it with information about successes or failures.” (Black et al 2003 p56)

This would seem to demonstrate that most teachers (even those teaching to the test) do want their learners to improve and that their teaching is tailored to this aim.

“The overall message is that summative tests should be, and should be seen to be, a positive part of the learning process.” (Black et al 2003 p56)

Since Black et al and their supporters excuse the vagueness of the supporting theory by insisting that the studies and resulting advice have been predominately empirically and practice based (see below in section 4) it perhaps is indicative of the importance of this deficit in strong, coherent theoretical framework that the teachers themselves, for whom the AfL paradigm has been formulated, have trouble putting its conclusions into practice.


In education, research pressure tends to divert energies to developing and examining classroom practice and providing evidence to support this: “the demand for evidence-based practice gives relatively little priority to theory” (Tight 2004 p406). Broadfoot and Black (2004 p10) come to essentially the same conclusion. To summarise, they acknowledge that their initial work on formative assessment had no firm theoretical basis but instead collected findings on research relating “to the notion” of formative assessment: the term notion is vague and unsubstantial.

“(Black and Wiliam 1998a, b) did not start from any pre-defined theoretical base but instead drew together a wide range of research findings relevant to the notion of formative assessment.” (Black and Wiliam 2009 p5)

This begs the question that if no theory was involved, how could they effectively and precisely selected relevant research for their review and paper. Indeed, Biggs 1998, in response to the review paper challenges the decision to exclude summative assessment from the choice.

Black and Wiliam 2009 summarise their past work and note their achievements and also signal certain deficits in theory. They conclude that Black et al. 2002, 2003 proffer practical advice and practices and “presented on a pragmatic basis, with a nascent but only vaguely outlined underlying unity” (Black and Wiliam 2009 p5): this work had drawn on Black and Wiliam 1998a, b which “did not start from any pre-defined theoretical base but instead drew together a wide range of research findings relevant to the notion of formative assessment” (Black and Wiliam 2009 p5). This reflects exactly Tight’s conclusions on research within education departments, which is, that evidence-based work has eclipsed considerations of theory.

Black and Wiliam 2006 “Developing a Theory of Formative Assessment” does not deliver on the expectation produced from the title: they situate their practice within frameworks of ‘pedagogy’ as recommended by Perrenoud (1998), but it does not develop the theory. The following citation confirms this and qualifies their focus:

“Subsequently, (Black and Wiliam 2006) we explored the changes that occurred in the classrooms of teachers developing formative assessment, and proposed a theoretical frame for the study of such classrooms. However, this theoretical frame was grounded in the data collected from classroom observations and interviews with teachers, and no systematic attempt was made to connect these data to work on such topics as classroom practice, or the regulation of learning.” (Black and Wiliam 2009 p6)

Again, as signalled by Tight, the principal priority is on classroom practice as opposed to pure theory of principles of assessment. The same can be said about subsequent work which is delineated in their summary:
“Other recent work has focused on aspects of implementation, notably on effecting change with communities of teachers (Wiliam 2007a) and on problems of superficial adoption (Black 2007), whilst both the book by Black et al. (2003) and the studies of the project on “Learning how to learn” (James et al. 2007) have discussed the learning principles underlying formative practices. In the conclusion of our 2006 article, we raised the wider issue of the role of formative assessment.” (Black and Wiliam 2009 p6)

In fact they seem to do everything but look at actual theory of either assessment or formative assessment and this is explicitly stated:

“So our first aim in this paper is to provide a unifying basis for the diverse practices which are said to be formative.” (Black and Wiliam 2009 p5)

If their aim is focused on ‘diverse practices’, why does their paper carry the misleading title of “Developing the theory of formative assessment”? In fact, why mention theory at all if this is irrelevant and it appears to be through out their work: irrelevant in as much as they pay it little attention, not that it lacks importance and priority. The use of the definite article in the title of the 2009 paper may simply be to distinguish it from the 2006 paper, or it may be implying a monopoly or claim on formative assessment theory, which like the 2006 it fails to deliver on.

7. Acknowledgement of the danger of theoretical deficit

As concerns theory, Stobart, a key member and supporter of the Assessment Reform Group and AfL, voices generally held beliefs:

“AfL is best viewed as an approach to classroom assessment to support learning, rather than as a tightly formulated theory”: the ‘spirit’ of AfL is important for understanding the “approach”. (Stobart 2008 p145)

After 30 years of AfL, and despite world-wide research on classroom practice it seems surprising that he believes that “…our understandings are still at an early stage” (Stobart 2008 p145). Stobart recognises the problems inherent in this. There is a danger that this may result in AfL being seen as “a series of classroom ‘handy hints’, rather than a theory-driven approach to teaching and learning” (Stobart 2008 p149). He further notes that

“This does not mean that there is no theoretical underpinnings; simply that it has not been organised, and may not need to be, into a stand-alone theory” (Stobart 2008 p145)

They seem to have taken theory as follows: the empirical work they examined had theoretical premises, therefore the results of this work had a theoretical basis, therefore collecting a number of these ideas and using them as interventions in the classroom meant that these interventions had a theoretical basis, therefore they had theory – a conceptual leap and non sequitur.

In fact feedback and FA are part of theory – but they have not linked definitions with theory, nor the fact that these definitions are required to be linked to process and then translated in practice, or at least linked to practice.

8. Proposing a theory of assessment: summative, formative and self-assessment

Taras 2005 uses definitions from Scriven 1967 and Sadler 1989 to explain and interlink the elements of the process of assessment. From Scriven (1967) there is the definition of assessment, and the original distinction between summative and formative assessment. Sadler’s theory of formative assessment and feedback (1989) provide an augmentation of Scriven’s argument. From these, extrapolations both from their work and ‘Taras’ own work on assessment and self-assessment provide the links and the relationships between the elements and the rationale which sustains them. A brief summary of this argument follows and the rationale which makes this a feasible and working theory to support the principles of assessment for learning.

Assessment is a judgement: parameters i.e. standards, criteria, context etc are required to make this judgement. These parameters may be implicit i.e. in the assessor’s head, or they may be explicit and shared with others. ‘This chair is comfortable because …’ is a judgement which shares some of the parameters with the listener. In an educational context, judgements may be of process, product or interactions. Therefore, everything can be assessed and this ubiquitous process is often taken for granted and rarely analysed or broken down.

However, Taras (2005) argues that making this explicit enables us to both understand, clarify and hence improve it. How the parameters of assessment are weighted and used can vary: Scriven’s (1967) definition requires the weighted criteria to be justified according to defined standards so the assessor communicates to others how, why and what priorities are attached to the assessment and the final judgement or grade. Sadler (1989), within the context of formative assessment in a multi-criterion academic piece of work, focuses on the importance of sharing and understanding the principal criteria in addition to being aware of implicit, basic criteria that only come into play
when they are not respected at the level within which learners are working. For example, the criterion of punctuation would not be used or made explicit at university level work unless it was not respected.

Both Scriven and Sadler require the sharing and understanding of the assessment parameters between the assessors and the assessed.

The result of a judgement or assessment may be represented in words as feedback, or in a summary judgement in the form of a grade or mark in accordance with an agreed scale. Feedback, according to Scriven’s definition of assessment, will justify the judgement: if it stops there and is a summation of work to date it is called summative assessment, if the feedback is used to update, change and improve the work, it becomes formative assessment. Sadler’s definition of formative feedback requires all feedback to be used in order for it to be called ‘feedback’: otherwise it is merely information produced by the assessor. Sadler requires the understanding of the parameters and information from the assessor as a condition for formative feedback.

Taras (2005) argues that all assessment, including formative assessment, requires and implicit or explicit judgement or summative in order to provide feedback which may subsequently be used as formative feedback and become formative assessment. She provides a formula to represent this:

\[ \text{SA} + \text{feedback} = \text{FA} \]

This framework provides definitions of the terms of assessment, how they inter-relate and the process which links them. It makes it clear that in this framework assessment is defined as a process, and does not add the confusion of functionality – the process of formative assessment can only be said to have taken place when feedback has been used to improve the work; whether this work is used for accreditation or within a less formal arena does not affect the definition. This is the theory which can explain and justify assessment for learning because it firstly, supports the principles, secondly, clarifies the process and thus permits it to be monitored and improved.

The above argument and rationale seems appropriate for discussions about product or process assessment of specific academic work. For the context of classroom interaction and learning which is the focus of AfL, Black et al 2003, proffer four interventions which formed the basis for AfL dissemination and support in schools, so it is appropriate to examine these in relation to the assessment framework presented above. However of these four interventions, three of them i.e. feedback through marking, the formative use of summative tests, and peer and self assessment would be product assessment and therefore appropriate for the above rationale. Questioning within classroom interaction should pose fewer problems for a formal assessment theory than Black and Wiliam would have us believe because this is in fact the ideal medium for negotiation of meaning within a social-constructivist theory of learning.

9. Conclusion

In this paper we have examined the claims of AfL in terms of theoretical consistency. It has become clear that even those internal to the AfL group have in hindsight lamented the deficit in theory. It has been shown that the theoretical problems make it difficult to apply its innovations in practice, as teachers (external to the AfL group) have found it impracticable. The problems themselves can be traced to a lack of clarity between the roles of function, product and process definitions of assessment, in a lack of transparency as to the relation between summative and formative assessment, and logical inconsistency within the idea of assessment as a process. It has never been in question that a solid backdrop of theory is essential to the working of any framework, and so the issues caused by vagueness in this arena is no surprise. However, the aims and ethos of AfL is valuable to the education community, and as such is deserving of a robust theory of assessment to support it. This can be found in the definitions of Sadler (1989), the work of Scriven (1967) and the theory of Taras (2005) which bring together all of the disparate ideas of formative and summative assessment into one cogent theoretical framework. It has been shown that this theory supports the work of Black and Wiliam and the AfL framework providing it with the substance it needs.

References


Black, P., & Wiliam, D, (1998b) Inside the Black Box: Raising Standards Through Classroom Assessment. PHI Delta Kappan, 80(2), 139-44.


