

## *The classroom assessment cycle within the alternative assessment paradigm: exploring the role of the teacher*

**Michael A. Buhagiar**

[michael.buhagiar@um.edu.mt](mailto:michael.buhagiar@um.edu.mt)

**Michael A. Buhagiar** is a lecturer of mathematics at the Junior College of the University of Malta and is an associate member of the Euro-Mediterranean Centre for Educational Research (EMCER) at the same university. His PhD (University of Nottingham, UK) focussed on sixth form mathematics teachers' classroom assessment practices. Dr Buhagiar is the managing editor of Mediterranean Journal of Educational Studies, the biannual international refereed journal published by EMCER.

### **Abstract:**

This review paper explores the role of the teacher in classroom assessment within the parameters set by the demands and expectations of the new, alternative assessment paradigm. After briefly outlining the underlying philosophy of this new paradigm, classroom assessment is presented as a cycle of four interrelated phases – namely, ‘planning the activity’, ‘gathering the evidence’, ‘interpreting the evidence’ and ‘using the evidence’. Within each phase, teachers’ classroom assessment practices are discussed in relation to how these compare with what is needed in order to bring assessment at the service of learning, which lies at the heart of our new understanding of assessment. The realisation that, generally speaking, teachers’ assessment practices remain firmly anchored to the traditional assessment theories and policies sends a clear signal that something needs to be done unless we want to risk reversing, with grave consequences for learning, the whole assessment reform process.

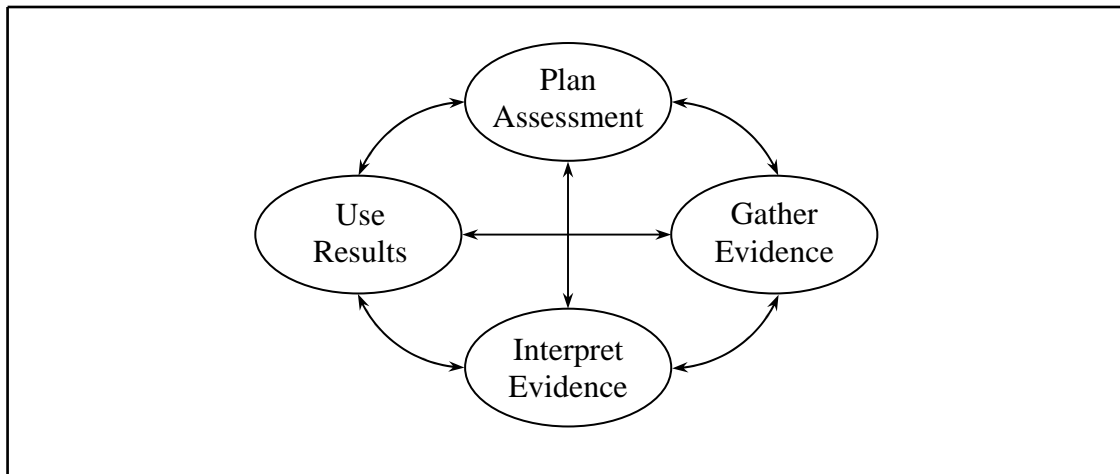
## **Introduction**

Contrary to the traditional paradigm in which assessment is seen purely in terms of its product, its results and how these results may be used to manage or even drive school systems, assessment in the alternative paradigm is seen as a process almost wholly integrated with teaching and learning (Torrance, 1995). I consequently find myself agreeing with Gipps (1994) when she argues that, according to our current conceptions of learning, of assessment and of what counts as achievement, the appropriate assessment model inside the classroom is one that is designed to support the teaching and learning of important skills and concepts. This newly emerging culture is now generally known, at least in the UK and subsequently also in Malta, as 'assessment for learning'. This basically embodies "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group [ARG], 2002).

Embedded within this vision lie significant changes in the traditional assessment roles of teachers and students. For just as teachers are no longer considered to be transmitters of knowledge but facilitator of student learning and the student is no longer considered to be a receiver but a constructor of knowledge, so too are they no longer expected to behave respectively in classroom assessment solely as the 'one who checks' and the 'one being checked'. The new paradigm calls in fact for classroom assessment to be seen as the gathering of information by both the teacher and students about their teaching-learning situation in order to help them in their decisions. It is now up to the teachers to understand their new role in classroom assessment and to work with students towards turning the new assessment philosophy into successful classroom realities. However, considering the deep rooting of the traditional assessment model and the accompanying largely 'not for learning' educational contexts in which teachers and students operate, this does not promise to be an easy journey.

## **The cyclic nature of classroom assessment**

Classroom assessment is normally presented as a cycle that is subdivided into a number of phases, often four (e.g., Bright & Joyner, 1998; Calfee & Masuda, 1997; Mavrommatis, 1997; National Council of Teachers of Mathematics [NCTM], 1995). Using NCTM's (1995) terminology, these are 'plan the assessment', 'gather evidence', 'interpret the evidence', and 'use the results'. This division is however arbitrary because "In practice, the phases are interactive, and the distinctions between them are blurred. Assessment does not proceed through them in a neat, linear fashion" (NCTM, 1995, p. 4; see Figure 1). Instead, all classroom assessment episodes occur within a sequence of interrelated phases (Mavrommatis, 1997) to ideally form a coherent whole. In the coming sections, my main concern is to explore these four phases from the perspective of what the teacher can do in order to better promote the 'assessment for learning' philosophy that is now widely accepted to represent the role of classroom assessment within the new assessment paradigm.

*FIGURE 1: The four phases of classroom assessment*

Reproduced from NCTM (1995, p. 4)

### **Planning the activity phase**

Although there is much to be gained for the teaching-learning process when teachers use formatively the information obtained from their ongoing informal assessment situations (i.e., which arises spontaneously from the naturally occurring classroom environment), one cannot simply let things ‘just happen’ for this would seriously jeopardise the quality of classroom assessment (see Bright & Joyner, 1998). In fact, given that classroom assessment, as it is now understood, is primarily about supporting learning, it is important for the teacher to gather as much ‘revealing’ information as possible, including through conscious planning, towards this end. The ultimate goal is for the teacher to have the necessary information to be able to plan work and to guide each student appropriately according to the learning goals of the course. Moreover, if students are truly to become insiders rather than consumers of classroom assessment (see Sadler, 1989), the teacher must find ways of involving them even at the planning stage. Within this new paradigmatic scenario, even the transfer of assessment information between teachers assumes a ‘for learning’ as opposed to a ‘reporting’ dimension (see Black, 1998). Another ‘planning’ consideration that is linked to the new paradigm is teacher collaboration. At issue is the need to substitute the traditional practice of teachers planning individually with the newer, open practice of teachers working as a group from the start, planning activities that would yield common assessment opportunities (see Torrance & Pryor, 1998).

Although I agree that teacher assessment plans should go well beyond the ‘next lesson’ context, I find Brookhart’s (1999) suggestion that these plans should be part of planning for a course from the very beginning a little unrealistic, especially in yearlong courses. My choice to favour a more flexible approach to assessment planning does not however preclude me from arguing like her that teachers need to prepare thoughtfully and carefully their assessment activities – both at the overall and the day-to-day levels – by answering the questions ‘What kind of information is needed?’ and ‘What performance by students will give that information?’ This

exercise, whose successful outcome largely depends on teachers having a clear sense of what they wish to assess and why they wish to assess it, should lay a solid foundation for the selection and use of proper assessment methods, which is an important prerequisite for quality assessment (see Stiggins, 1992). The emphasis in the new paradigm on linking assessment to learning does not however exclude that teachers, on certain occasions, other than using tasks for either diagnostic (before) or embedded (during) assessment purposes, also use tasks for mastery (final) assessment purposes (see Bryant & Driscoll, 1998). This is in line with the understanding that classroom assessment, apart from providing the information that is needed for immediate short-term purposes, is also used for summative long-term decisions (see Calfee & Masuda, 1997). For even when, as is mostly the case in Malta, assessments by teachers are not used for external purposes, the school itself is likely to want them to generate assessment information for internal purposes (e.g., streaming, promotion and reporting to parents).

### **Gathering the evidence phase**

In the new paradigm, the evidence-gathering phase is about gathering adequate and relevant information about students' learning. The idea is to obtain as comprehensive a picture as possible about the teaching-learning situation. This calls for a gathering approach that, apart from tapping evidence from a variety of sources that are either pre-planned or that arise spontaneously during the lesson (see Airasian, 2000), is also guided by the aim of primarily seeking evidence that illuminates each student's learning trajectory as opposed to comparing him or her against other students or norms. The evidence-gathering procedures that concern me here are directly related to the classroom situation – these include observation of students, written and oral communication, assessment tasks and class tests. But the teacher may begin gathering information about students, both at individual and group levels, before even meeting them. For teachers are known to have their 'antennae' up at the start of school, constantly searching their environment for information about students – an exercise that Airasian (2000) calls 'sizing up assessment'. Teachers obtain the outside classroom component of this 'sizing up' information from sources like the school grapevine, comments by other teachers, school records and performance of siblings.

#### *(i) Observation*

Much of the inside classroom component of this sizing up assessment comes from the informal observation of momentary unplanned happenings, such as when a student does or says something, that the teacher mentally records and interprets (Airasian, 2000). The teacher uses this information, together with what is learned from outside classroom sources, to form an initial set of perceptions and expectations about students that will then influence the way in which he or she plans for, interacts with, and manages students and instruction (Airasian, 2000). My main concern with this reality is that, as Airasian (2000) points out, these early impressions – about whose accuracy teachers are generally very confident – tend to become permanent, virtually stable throughout the year. This means that the teacher often forms generalised and lasting impressions from early singular or limited instances, practically from what the student happens to be doing or saying when the teacher glances his or her way in the first few days or weeks. The unfortunate fact that assessment so used prejudices rather than aids learning leads in turn to what I consider as an even graver concern – namely,

the stereotyping and labelling of students. For this triggers in students the mechanisms of the self-fulfilling prophecy that sees them going on to produce a reality that reflects these original evaluations by the teacher (see Filer, 2000).

Informal observation is however an important feature of classroom assessment right through the year, not only when teachers engage in their early sizing up assessment. Teachers do in fact rely heavily on observation to assess instruction (Airasian, 2000; Stiggins & Bridgeford, 1985). In particular, should their priority be to keep the teaching process going, they use the reaction of pupils to judge whether it is feasible to carry on (Black, 1998). Although, given the fast pace of classroom activity, it is quite understandable why, as Airasian (2000) contends, the primary indicators that teachers use to monitor instruction are those that are most readily available, most quickly surveyed, and least intrusive – basically, reactions from students such as facial expressions, posture, participation, questions, and attending – I would still argue like him that such reactions do not provide direct evidence of student learning, which is the real criterion of instructional success. What makes the situation even less tenable is the tendency of teachers, irrespective of whether this arises from seating arrangements or from their unconscious preference for certain students, to often focus on an overly narrow sample of students and being inattentive to the rest (see Airasian, 2000). This facet of classroom assessment – which is characterised by the teacher's probable impossibility to monitor the classroom experience of each student whilst fully engaged with instruction – is not conducive to the rich, individualised information that is needed to help students progress in their learning (see Calfee & Masuda, 1997).

This realisation does not however mean that there should be no place for informal observation in classroom assessment. For, in truth, even facial expressions may reveal relevant diagnostic information (Broadfoot, 1996). My position is consequently that since such observations produce evidence, however ephemeral, that has its own unique value, it makes good sense to continue with this evidence-gathering procedure. This can only be problematic 'for learning' should the teacher fail to collect also corroborative evidence through the use of multiple methods that guarantee higher quality (as the strengths in one source compensate for the weaknesses in others) and fairer information (as all assessment methods may be said to have a certain amount of bias). Other than this, the quality of observational evidence itself can be enriched should the teacher give, even if only at times, a more formal and planned dimension to his or her observations. Some studies show in fact that teachers find it surprisingly useful to suspend their active teaching – making clear to the class what they are doing and why – and to concentrate only on looking and listening with a few students at a time whilst the rest are engaged in individual or small-group activities (Black, 1998). This is in line with Calfee and Masuda's (1997) assertion that assessment through observation improves when teachers create specific occasions for observation and practise 'focus' – that is, they select what and whom to observe, and put all else in the background.

#### *(ii) Communication*

Questioning is the most common form of evidence-gathering technique used by teachers when acting deliberately to obtain information about students' knowledge or capabilities (William & Black, 1996). But albeit the teacher can question students both

orally and in writing, it is the oral form that is normally practised. The popularity of oral questioning emerges from research showing that on average teachers may spend almost half their time on it (see Broadfoot, 1996). The almost parallel exclusion of written communication – which may take the form of the ‘pupil journals’ mentioned by Stiggins (1997) – works especially against teachers getting to know those students who, even when justly so are not publicly grilled through questioning, are ‘tongue-tied’ in classroom discussions and can express themselves better in writing. Whilst these missed ‘written opportunities’ somewhat undermine the rich data oriented spirit of the new paradigm, it must also be said that, as Brookhart (1999) points out, although questions in class help both teachers and students to clarify what students know and where their misconceptions have occurred, giving an accurate indication in the process of what the class as a whole understands, the information obtained from them does not give a complete picture of an individual’s understandings. In truth, not only is there no guarantee that questioning elicits if a student has any particular knowledge or understanding, but one can also never be sure that a student does not know something through questioning (Wiliam & Black, 1996). I would consequently argue like Pryor and Torrance (2000) that teachers would do well not to treat the answers to their questions as unproblematic sources of information for pedagogic decision-making.

The need for teachers to reflect on the information received through questioning is also linked to what the different types of teacher questions – which are usually classified along a variety of interrelated categories – can possibly reveal about and contribute towards students’ learning. With regards to categories, one can speak for instance of convergent and divergent questions and of higher and lower level questions (see Airasian, 2000). Convergent questions (also called ‘closed’) have a single correct answer and divergent questions (also called ‘open’) may have many appropriate answers. On the other hand, whilst lower level questions require students to simply retrieve and manipulate factual knowledge, higher level questions require students to build on this factual recall and engage in solving new problems. Although I agree with Airasian (2000) that there is a place during instruction for these different kinds of questions – suffices the knowledge that factual recall is the basis for higher level questioning – my position is that one needs however to prioritise amongst them. In particular, rather than emphasise closed questions that possibly leave students calculating whether to take the risk on their chances of knowing the right answer, teachers should invest more in open questions that show students that the teacher is interested in their ideas, encourage students’ self-expression and challenge students to develop their thinking (see Black, 1998). Such a line of questioning that favours the creation of a classroom environment that is open to the potential of discussion as a learning tool (see Swan, 2001) parallels Torrance and Pryor’s (1998) advocacy for the use of ‘genuine questions’. Their argument is that unless questions elicit ‘genuine’ or ‘authentic’ responses from students (as opposed to prefabricated responses presumed by students to be what teachers want to hear), they would be grounded in the exigencies of teaching (i.e., to move the lesson forward – see also Broadfoot, 1996) rather than the promotion of learning. On the contrary, genuine questions, apart from providing insight into students’ current state of understanding, are also potentially useful in stimulating further learning (see Torrance & Pryor, 1998).

But in contrast to this formative promise, Black and Wiliam (1998) conclude from their extensive review of the literature that the quality of classroom questioning is a

matter of concern. Not only is questioning at all classroom levels dominated by recall questions (Stiggins et al., 1989) that follow the traditional ritual sequence of ‘question by teacher, response by student(s), and feedback/evaluation by teacher’, but teachers often also choose a sub-group of only a few students, and it is their reactions and responses to questions which serve to justify proceeding (Black, 1998). The overwhelming quantity of talk during classroom discussion moreover comes from the teacher, with very few words being actually spoken by students (Torrance & Pryor, 1998). This reflects a reality in which it is more common for a teacher’s goal to be simply that of eliciting the correct answer from students rather than to engage them in discourse that requires them to articulate, develop and defend positions (see Calfee & Masuda, 1997). Teachers are so seduced into seeking and hearing correct answers, which then enables them to make a favourable judgement about their instruction, that they prefer factual questions to open-ended, complex ones in order to ensure more student participation and mastery (Airasian, 2000). This seduction is such that when they ask a question to students, they just move quickly around the class until they hear the right answer – conveying in the process an impression to students that it is speed that is important rather than thinking deeply about things (Boaler, 1997). The point is that questioning inhibits rather than helps the learning process when, as Lesh et al. (1992) contend, students are probably passed over by the teacher if they take more than three seconds to respond. Especially when faced with higher level questions, students need time to process their thinking in order to come up with more complete, thoughtful responses (Airasian, 2000). This calls for teachers to provide ample time and then to listen sensitively so that they pick up clues about a student’s thinking that might need to be followed up (Black, 1998).

### *(iii) Tasks*

For teachers not to rely unduly on the ephemeral evidence of classroom events, students need to systematically produce written work both in class and at home (Black, 1998). This calls, however, for assessment tasks that work towards valued learning goals and that are open in their structure to the generation and display of relevant evidence to the teacher and to the students themselves (Black & Wiliam, 1998). Towards this end, the teacher should select tasks that are “novel and varied in interest, offer reasonable challenge, help students develop short-term self-referenced goals, focus on meaningful aspects of learning and support the development and use of effective learning strategies” (Black & Wiliam, 1998, p. 31). The tasks used should reflect current learning theories that configure the teacher’s role as that of helping students find, create and negotiate their meanings by providing them with meaningful and purposeful activities from their perspective (see Murphy, 1996). This positioning heavily curtails the use of atomised assessments – that is, when specific skills are assessed out-of-context rather than as part of a realistically complex task (see Black, 1998) – that reveal very little about students’ thinking. At issue is the need for classroom tasks to be ‘authentic’ so as to facilitate the development of students’ understandings into knowledge that can be applied in real-life contexts, thus ensuring an explicit link between school learning and out-of-school practices (Murphy, 1996). These are tasks that, as Eisner (1993) points out, should:

- reflect the tasks that students will encounter both inside and outside schools;
- reveal how students go about solving a problem, not only the solutions they formulate;

- reflect the values of the intellectual community from which the tasks are derived;
- not be limited to solo performance;
- make possible more than one acceptable solution to a problem and more than one acceptable answer to a question;
- have curricular relevance, but not be limited to the curriculum as taught;
- require students to display a sensitivity to configurations or wholes, not simply to discrete elements;
- permit the student to select a form of representation to display what has been learned.

I see in this move away from the traditional decontextualised, rote-oriented tasks that impose low cognitive demands on students a shift towards a form of instruction that emphasises meaningful learning (see Darling-Hammond, 1994). Given that tasks constitute key contexts for students' thinking about the subject (Doyle, 1988), it follows moreover that the teacher, in his or her role of task selector, needs to possess a skilled and multi-dimensional foresight (Black, 1998). For not only must he or she reckon with constraints of time, of facilities and of the starting-point of the students (Black, 1998), but attention must also be paid to the content of tasks as this sends a clear message to students about what parts of the subject are important to learn (Bryant & Driscoll, 1998) and the manner in which students are expected to work on the tasks as this delineates their learning habits. Our present understanding of the learning process makes it vital for students to be involved in collaborative projects, as these create the conditions for thinking aloud and sharing ideas, which is an important metacognitive aspect of learning and assessment so often lacking when students work alone in traditional school assessments (Ellis, 2001). Unfortunately, group work remains shunned by some teachers because they prefer to do all the talking themselves, by others because they prefer the silent atmosphere of a classroom where each student is busy doing his or her 'own' work, by others because they fear that this would limit the amount of work they can cover in a lesson (Ellis, 2001). And still by others because they have problems with student motivation, or what has been called 'free riding', which is a form of social loafing seen in a group when one or more members slack off and 'ride' on the extra efforts of their coworkers (Walker & Angelo, 1998).

Another important consideration in task selection that teachers have to grapple with is the degree to which the task is left open or closed. Whilst closed tasks are linked to standard textbook questions, school-learned methods and rules (i.e., tasks that encourage the development of procedural knowledge in students), open tasks are linked to practical and investigative work that requires students to make their own decisions, plan their own routes through tasks, choose methods, and apply their knowledge (i.e., tasks that encourage the development of conceptual knowledge) (Boaler, 1998). Apart from such considerations, a task can also be specified according to the complexity of reasoning it requires. Black and Wiliam (1998) refer in fact to a scheme developed by Dumas-Carre and Larcher in 1987 that can be used to produce such a comparative and descriptive analysis of tasks:

This scheme distinguished tasks which (a) presented a specific situation identical to the one studied, or (b) presented a 'typical' problem but not one identical to the one studied, requiring identification of the appropriate algorithm and its use, rather than exact replication of an earlier procedure



as in (a), and (c) a quite new problem requiring new reasoning and construction of a new approach, deploying established knowledge in a new way. (pp. 31-32)

Clearly, as one moves from (a) to (c), the level of student thinking involved in working with tasks evolves from lower level (characterised by mere recall of factual information) to higher level (characterised by the application, analysis and synthesis of factual knowledge in order to solve new problems). But although teachers have such a wide array of tasks at their disposition – both with regards to openness and complexity – it is as if the thinking level demanded by a task is inversely proportional to its classroom use because, as Carter and Doyle (1987) point out, higher order tasks are rarely given in class. And when potentially demanding tasks are set, teachers avoid classroom conflicts by ‘redefining or simplifying task demands’ (Doyle, 1988). This teacher reluctance to spend time on what are basically nonroutine activities characterised by conceptual understanding, explorations, construction of meanings and invention – which is in direct conflict with the learning demands of the new paradigm – results from their perception (which is often correct) that these are irrelevant to students’ examinations (Goldin, 1992).

#### *(iv) Class tests*

Testing remains synonymous with schooling (see Ellis, 2001). Not only do the majority of teachers spend more than 10% of their professional time on testing (Newman & Stallings, 1982; cited in Schafer, 1993), but teachers are also inclined to use tests irrespective of the purpose of assessment (Stiggins & Bridgeford, 1985). This reality persists even if, as Gipps and Murphy (1994) argue, fair tests as such do not and cannot exist. It consequently makes no sense to expect tests to establish and to provide accurate feedback about what the student actually knows at a particular point in time (see Torrance & Pryor, 1998; also Gipps & Murphy, 1994). Apart from the test itself, its context may also make a significant difference. For instance, a low stakes test situation is unlikely to draw forth highly motivated best performances, which means that the data derived from such an exercise may not constitute a particularly valid indicator of educational achievement (Torrance, 1995).

In spite of its challenge to the traditional behaviourist rhetoric, the new assessment paradigm reconceptualises rather than abolishes the use of tests as evidence-gathering instruments. Constructivist theories demand in fact that tests show what students know and can do, as well as facilitate good learning – what Glaser (1990) calls ‘placing tests in the service of learning’. Within this emerging framework, tests should consequently be “ambitious instruments aimed at detecting what mental representations students hold of important ideas and what facility students have in bringing these understandings to bear in solving their problems” (Shepard, 1991, p. 9). The new emphasis on integrating tests to instruction in order to render them useful for instructional decisions (Black, 1998) builds on the understanding that it is the manner in which test results are interpreted and used by teachers and students alike that determines whether or not testing actually serves the formative or the summative function.

Notwithstanding their formative potential, I remain concerned with what lies behind the continued proliferation of testing inside the classroom. It is, for instance, worrying that whilst teachers are concerned about the time required to develop and use their own tests as this interferes with their instructional time, they tend to be less concerned about their lack of information on testing, their competence in testing, the student reaction to testing, and collaborating with others in testing (see Stiggins & Bridgeford, 1985). Probably of more concern is teachers' tendency to make little use of tests results beyond putting them into record books and using them to identify students for remedial help (Gipps et al., 1995). This little use of test results – which arises from teachers' inability to see tests as saying something about their teaching rather than just about the student (Wood, 1990) – is in line with teachers' general unwillingness to adapt the curriculum in response to testing (see Close & Brown, 1987; cited in Gipps et al., 1995). The 'unhealthy' distancing between testing and instruction is again evident in the manner in which class tests are designed to mimic the examinations used to certify achievement. This reproduces at classroom level the same problems that are generally associated with such examinations (e.g., the message that the rapid use of well-learned techniques is most important). The shortsightedness of this mimicry emerges also from the studies that show how class tests can improve student examination performance without any real or lasting improvement in educational quality (see Torrance, 1995; also Shepard, 2001 [cited in Schoenfeld, 2002]).

### **Interpreting the evidence phase**

Collected evidence needs to be interpreted so that it may be turned into information on the basis of which decisions can be made (see Wiliam & Black, 1996; also Calfee & Masuda, 1997). As far as the teacher is concerned, the examination of the evidence helps him or her to determine whether or not there is a gap between what students can actually do and what he or she would like them to be able to do (Wiliam & Black, 1996). This reflective exercise helps

... teachers decide if instruction is being effective so that changes and modifications can be made. ... Some of the reflections will be formative 'evaluation' of students' progress, and some will be summative 'evaluation' that compares students' progress against established standards of performance. (Bright & Joyner, 1998, p. 31)

Although I hold that teachers should retain their summative role as this can benefit the quality of summative assessments through the inclusion of skills, competencies and knowledge that cannot be assessed by the more traditional paper-and-pencil approach (see Broadfoot, 1996; Broadfoot & Black, 2004), I would still argue from an 'assessment for learning' perspective that the standards against which they compare the evidence should be primarily self-referenced or at least criterion-referenced, not norm-referenced (see Mavrommatis, 1997). This does not however exclude that the teacher, apart from interpreting the assessment data from a singular frame of reference in order to make decisions about single students, also views the data from a collective frame of reference in order to make group instructional decisions (see Phye, 1997).

In either case, for truly professional judgements, teachers need both time and occasion to think about the evidence, ideally in consultation with colleagues (Calfee & Masuda, 1997). For even if the interpretation of evidence is typically tacit and intuitive, based

upon knowledge of students that teachers would have acquired through experience at both the collective and individual levels (Mavrommatis, 1997; Watson, 2001), it has to be said that teachers are so assailed by information in the classroom from all sides that they rarely have the time to make considered decisions in the moment (Watson, 2001). It is again limiting on the quality of the interpretations made – and consequently on students' learning – that teachers are often, according to Wiliam and Black (1996), the sole interpreters inside the classroom of the assessment evidence. Students' involvement is essential because, apart from adding a different informed perspective to that of the teacher, they need to come to understand their strengths and weaknesses, and how they may deal with them (see Harlen & James, 1997) if they are to become self-monitoring learners.

### **Using the evidence phase**

The interpretations that teachers give to assessment results are a means to an end. In fact, with very few exceptions, assessments are conducted for a purpose and certain actions follow the outcomes (Wiliam & Black, 1996). These actions or consequences can be grouped under three interrelated categories – namely, instructional decisions (which relate to the teacher interventions aimed at improving learning), feedback (which is mainly related to the formative function of assessment) and grading (which is primarily summative in nature). But given that these actions target different audiences, the need arises for teachers to have an adequate recording system in order to be able to select, edit and communicate assessment information appropriately and effectively. I find that Murphy and Torrance's (1988) distinction between formative and summative recording provides the framework on which teachers can build such a system:

Formative records are essentially internal working documents, continuously updated and amended, for use by both teacher and pupil to encourage, guide and reward learning and to stimulate reviews of the curriculum and pedagogy by informing teachers of the effectiveness of teaching methods and the appropriateness of what is taught. Summative records are static, end-of-stage ... documents which present a distillation of all the assessment information available about a pupil geared, both in terms of content and format, to the needs or interests of audiences outside the school ... (p. 63)

#### *(i) Instructional decisions*

Teachers continually make instructional decisions according to their knowledge of what students know and can do (Bright & Joyner, 1998). Many of these decisions – which may involve proper instructional interventions or revision of tasks and assessments – are actually taken during the course of instruction itself on the basis of how teachers interpret ongoing assessment evidence (Airasian, 2000). In these circumstances, the teacher has to decide there and then whether the lesson is progressing satisfactorily (in which case, the lesson continues according to plan) or whether a problem is sensed (in which case, the teacher either revises the planned instructional activity or initiates another teaching activity) – a teaching-assessment cycle that is repeated many times in the course of a single lesson (Airasian, 2000). The problem here is that it is hardly ever feasible for the teacher to monitor in detail

the progress of each individual student during instruction. It is far more likely that the teacher monitors the impact of his or her teaching at an overall level rather than at the individual student level (Torrance & Pryor, 1998). Such a reality, characterised by the teacher focusing on individuals in detail only if they are causing real concern (Torrance & Pryor, 1998), unavoidable in practice as it may be, works against the realisation of each student's learning potential, which lies at the heart of the new paradigm. To act formatively, the teacher needs instead to have detailed, quality information about individual students. For only then can he or she have the opportunity to put students in learning situations that are potentially optimal for them, and to optimise the activity and the learning process of each student within a given situation (see Perrenoud, 1998). It thus makes sense for the teacher to delay taking important instructional decisions, possibly only acting in subsequent lessons, until such information is available. By avoiding to base decisions on biased evidence, the teacher would be lessening the chances of producing invalid conclusions about the success of instruction with harmful consequences for students (see Airasian, 2000).

### *(ii) Feedback*

In the new paradigm, the basic issue with feedback is that it should provide learners with constructive guidance about how to improve (ARG, 2002). Feedback is thus about the promotion of a culture of success where students can build achievements on their previous performance without any comparison with others (Black et al., 2003). This understanding is in line with Ramaprasad's (1983) argument that feedback is actually feedback only if it satisfies the basic condition laid down in his definition:

Feedback is the information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way. (p. 4)

According to this definition, should someone discover that there is a gap but still has no idea about the nature of the discrepancy between actual and desired performance, then that information – which is almost inevitably norm-referenced – fails to qualify as feedback as it does not help him or her to close the gap. Such a process is better described as simply 'monitoring' (William & Black, 1996). On the contrary, for assessment information to count as feedback, it must indicate the existence of a gap between actual and desired levels of performance, as well as suggest actions that prove successful in closing the gap (William & Black, 1996; also Black et al., 2003). This requirement, which puts feedback at the service of learning, is linked in turn to the concept of student agency as a part of having each student become an independent learner. Building on Ramaprasad's notion of feedback, Sadler (1989) argues in fact that in order for the student to improve, he or she must: (i) have a notion of the desired standard or goal; (ii) be able to compare the actual performance with the desired performance; and (iii) engage in appropriate action to close the gap between the two. This can only happen, however, if teacher's standards are available to the student and teacher feedback allows the student to reach these standards (Sadler, 1989). The embedded understanding that feedback should constitute episodes of learning that enable students to connect aspects of poor performance to specific remedial actions encourages me to argue, like Stefani (1998), that what students require is user-friendly information that relates to how they are doing and how specifically they might be able

to improve upon this. In this respect, Wiggins (1993; cited in Stefani, 1998) provides teachers with some valid indicators of what it might mean to provide good feedback:

- define the requirements of each learning task;
- describe clearly how performance will be measured/graded/assessed, preferably involving students in this process;
- provide well-articulated descriptors or exemplars of different levels of attainment;
- provide feedback about individual performance expressing this in accordance with agreed criteria;
- relate various aspects of poor performance to specific remedial actions.

These indicators are as much about the communication of expectations (i.e., feedforward) as they are about the communication of progress towards goals (i.e., feedback). In either case, the emphasis is on all students (as opposed to the current practice of having the better students receiving more feedback) receiving feedback that has high-communication value, in the sense that it can be understood and used by them (Stiggins & Conklin, 1992). This clearly excludes, in spite of many teachers still believing otherwise, that a grade and a short series of comments, usually of a simple praise or blame nature, constitute feedback (Stefani, 1998). On the other hand, if feedback is to provide teachers, as it should, the 'vehicle for personal dialogue with each learner' (see Black et al., 2003), it must be given regularly and whilst still relevant, and should also focus on and be task specific (Crooks, 1988). In particular, given that students' self-perception as learners depends on the quality of feedback they experience over time (Black, 1998), it is essential for feedback to direct attention to the task rather than the learner, as this would lessen the likelihood that the less successful students see it as another confirmation of their inability to perform, yet a further blow to their already low self-esteem (see Wiliam, 1998; also ARG, 2002).

### *(iii) Grading*

According to Harlen et al. (1992), there are two main ways in which teachers produce summative information about students. These are 'summing up' and 'checking up':

The former is some form of summary of information obtained through recording formative assessments during a particular period of time and the latter the collection of new information about what the pupil can do at the end of a period of time, usually through giving some form of test. (p. 222)

The possibility that teachers 'sum up' formative information for summative purposes – which indicates that these two forms of assessment are not mutually exclusive (see Torrance & Pryor, 1998) – adds credibility, I find, to my position that, within the new paradigm, it is not incompatible for teachers in systems dominated by assessment for selection and certification purposes to have a summative role. For me, instead, the real issue in similar circumstances is to ensure quality in the summative reporting procedures. Given that, for the foreseeable future, teachers will continue, at the most formal level, to judge and to communicate information about student performance through grading (see Airasian, 2000; Brookhart, 1999), my plea for quality summative reporting is basically, at least for the time being, a call for quality grading. My position is that, as long as teachers are required to grade students, we must make sure that grades carry 'real meaning' and be appropriate for the purposes to which the

users of their information will put them (see Brookhart, 1999). Airasian (2000) identifies four such purposes: (i) *administrative* – schools need grades to determine things such as suitability for promotion; (ii) *informational* – grades are used to tell parents, students, and others about a student's academic performance; (iii) *motivational* – the promise of a high grade is used to motivate students to study; and (iv) *guidance* – grades are used to guide students and parents choose appropriate courses and course levels, and then by schools to sanction or veto these choices.

Although these purposes can be satisfied, almost invariably better, by alternative assessment means, such as profiling and portfolios, that are more conducive to learning than grading, there is little doubt that these new approaches have so far failed to leave a lasting impact on most educational systems (see, for example, Grima & Chetcuti, 2003; Murphy & Torrance, 1988; Weeden et al., 2002). Until such time, there is however an urgent need to improve grading in order to protect the classroom-learning environment. I say this in the knowledge that grades tend to encourage cheating and can negatively influence students' motivation and self-esteem when lower than expected or consistently low (see Airasian, 2000). Grades may also reward rote learning and foster competitive and grade-hunting attitudes (see Mavrommatis, 1997). Moreover, when grading is cumulative as part of a continuous assessment system, such as when each attempt or piece of work submitted by a student is scored and the scores are added together at the end of the course, students may develop a 'not for learning' mindset that it is only worth doing work that contributes to the total (see Sadler, 1989).

Loyd and Loyd's (1997) four grading principles offer, in my view, a sense of direction towards an enhanced grading process. These are: (i) the grading system should be clear and understandable; (ii) the grading system should be communicated to all stakeholders; (iii) grading should be fair to all students; and (iv) grading should support, enhance, and inform the instructional process. Things are however unlikely to improve unless, contrary to what happens at present, all teachers start getting formal training in grading and are provided with proper guidance about grading policies and expectations (see Airasian, 2000). This development may help, for instance, to change the practice, reported in a number of studies, of having grading being almost solely based on academic evidence of student achievement, with non-academic evidence (e.g., effort and improvement) as a basis for adjustment in student grades, not as the central determiner of grades (see Airasian, 2000; also Buhagiar, 2005). Although this probably reflects teachers' preoccupation with assigning marks that are publicly defensible (see Peterson & Stack, 1998), it neither does justice to the complexity of the processes involved nor does it lend itself to support learning. On the other hand, useful grading not only draws on several different types of relevant and valid information that gives students more opportunity to show what they can do (Airasian, 2000), but is also accompanied by specific teacher comments about the strengths and weaknesses of a student's work (Mavrommatis, 1997). By adding his or her interpretation to the formal reporting of results, the teacher can put the results in context, identify progress, explain difficulties and indicate ways in which fellow teachers, students, parents and employers can use the information creatively and maximally (Eggleston, 1991).

### **Lessons learnt**

I pointed out in the Introduction that it is not an easy endeavour to translate the spirit of the new assessment paradigm into standard classroom practices. In fact, the reality presented here shows that whilst educational theories and policies are forcefully pushing towards ‘assessment for learning’, the classroom realities in many countries, including the US and the UK which have been at the forefront of the assessment reform efforts, remain dominated by practices ingrained in the traditional ‘assessment of learning’ for the purposes of grading and reporting that has its own well-established procedures (see ARG, 1999). Not surprisingly, Malta – which is, educationally speaking, a satellite country – is experiencing the same implementation difficulties in spite of our many efforts and good intentions (see Buhagiar, 2005; Grima & Chetcuti, 2003). For reasons that go beyond the scope of this paper (I argue elsewhere that improvement in classroom assessment calls for action at the teacher, school and national levels, all of which work interactively – see Buhagiar, 2005), teachers appear unprepared rather than unwilling to take up the challenges of the new assessment paradigm.

These challenges are indeed real and tough; no amount of whitewash can ever turn them into opportunities. In particular, given that classroom assessment is a cycle of phases, it is enough to have one weak phase to possibly jeopardise the success of the whole process. The still unsatisfactory situation depicted in this paper signals unequivocally that policies by themselves, however good, do not automatically translate into the intended practices. As a matter of fact, the more successful countries – Australia is a case in point (see Butler, 1995) – have had the foresight to construct around their policies an all-encompassing ambience that helps assessment truly become an integral element of the learning process. This is, I believe, the way forward if we truly want the assessment reform process to go ahead in spite of the persistent calls from some ‘interested’ individuals to turn back the clock to the ‘good old days’ when, in reality, assessment was used primarily to motivate and push the few at the expense of the rest.

### **Acknowledgements**

Thanks are due to Mauro Scerri for reproducing the diagram in Figure 1.

## References

- Airasian, P. W. (2000). *Assessment in the classroom: a concise approach* (2<sup>nd</sup> ed.). Boston: McGraw-Hill.
- Assessment Reform Group (ARG) (1999). *Assessment for learning: beyond the black box*. Cambridge: School of Education, University of Cambridge.
- Assessment Reform Group (ARG) (2002). *Assessment for learning: 10 principles* [Leaflet/poster].
- Black, P. (1998). *Testing: friend or foe? The theory and practice of assessment and testing*. London: The Falmer Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74.
- Boaler, J. (1997). *Experiencing school mathematics: teaching styles, sex and setting*. Buckingham: Open University Press.
- Boaler, J. (1998). Open and closed mathematics: student experiences and understandings. *Journal for Research in Mathematics Education*, 29(1), 41-62.
- Bright, G. W., & Joyner, J. M. (1998). Understanding and improving classroom assessment: summary of issues raised. In G. W. Bright & J. M. Joyner (Eds.), *Classroom assessment in mathematics: views from a National Science Foundation working conference* (pp. 27-57). Lanham, MD: University Press of America.
- Broadfoot, P. M. (1996). *Education, assessment and society: a sociological analysis*. Buckingham: Open University Press.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of 'Assessment in Education'. *Assessment in Education: Principles, Policy and Practice*, 11(1), 7-27.
- Brookhart, S. M. (1999). *The art and science of classroom assessment: the missing part of the pedagogy*. Washington, DC: The George Washington University.
- Bryant, D., & Driscoll, M. (1998). *Exploring classroom assessment in mathematics: a guide for professional development*. Reston, VA: NCTM.
- Buhagiar, M. A. (2005). *Mathematics teachers' classroom assessment practices: a case study in a Maltese sixth form college*. PhD thesis, School of Education, University of Nottingham, UK.
- Butler, J. (1995). Teachers judging standards in senior science subjects: fifteen years of the Queensland experiment. *Studies in Science Education*, 26, 135-157.



- Calfee, R. C., & Masuda, W. V. (1997). Classroom assessment as inquiry. In G. D. Phe (Ed.), *Handbook of classroom assessment: learning, adjustment, and achievement* (pp. 69-102). San Diego, CA: Academic Press.
- Carter, K., & Doyle, W. (1987). Teachers' knowledge structures and the comprehension processes. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 147-160). London: Cassell.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Doyle, W. (1988). Work in mathematics classes: the context of students' thinking during instruction. *Educational Psychologist*, 23(2), 167-180.
- Eggleston, J. (1991). Teaching teachers to assess. *European Journal of Education*, 26(3), 231-237.
- Eisner, E. W. (1993). Reshaping assessment in education: some criteria in search of practice. *Journal of Curriculum Studies*, 25(3), 219-233.
- Ellis, A. K. (2001). *Teaching, learning and assessment together: the reflective classroom*. Larchmont, NY: Eye On Education.
- Filer, A. (2000). Classroom contexts of assessment: editor's introduction. In A. Filer (Ed.), *Assessment: social practice and social product* (pp. 83-86). London: RoutledgeFalmer.
- Gipps, C. V. (1994). *Beyond testing: towards a theory of educational assessment*. London: RoutledgeFalmer.
- Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence?* Buckingham: Open University Press.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, 14(5), 475-483.
- Goldin, G. A. (1992). Toward an assessment framework for school mathematics. In R. Lesh & S. J. Lamon (Eds.), *Assessment of authentic performance in school mathematics* (pp. 63-88). Washington, DC: American Association for the Advancement of Science Press.
- Grima, G., & Chetcuti, D. (2003). Current assessment practices in schools in Malta and Gozo: a research report. *Journal of Maltese Education Research*, 1(2), 57-94. Available online at: <http://www.educ.um.edu.mt/jmer>

- Harlen, W., Gipps, C., Broadfoot, P., & Nuttall, D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, 3(3), 215-230.
- Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy and Practice*, 4(3), 365-379.
- Lesh, R., Lamon, S. J., Behr, M., & Lester, F. (1992). Future directions for mathematics assessment. In R. Lesh & S. J. Lamon (Eds.), *Assessment of authentic performance in school mathematics* (pp. 379-425). Washington, DC: American Association for the Advancement of Science Press.
- Loyd, B. H., & Loyd, D. E. (1997). Kindergarten through grade 12 standards: a philosophy of grading. In G. D. Phye (Ed.), *Handbook of classroom assessment: learning, adjustment, and achievement* (pp. 481-489). San Diego, CA: Academic Press.
- Mavrommatis, Y. (1997). Understanding assessment in the classroom: phases of the assessment process – the assessment episode. *Assessment in Education: Principles, Policy and Practice*, 4(3), 381-399.
- Murphy, P. (1996). Defining pedagogy. In P. F. Murphy & C. V. Gipps (Eds.), *Equity in the classroom: towards effective pedagogy for girls and boys* (pp. 9-22). London: The Falmer Press.
- Murphy, R., & Torrance, H. (1988). *The changing face of educational assessment*. Milton Keynes: Open University Press.
- National Council of Teachers of Mathematics (NCTM) (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes: towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice*, 5(1), 85-102.
- Peterson, J., & Stack, C. (1998). A Minnesota story: a system approach to Classroom Assessment and Research. In T. Angelo (Ed.), *Classroom Assessment and Research: an update on uses, approaches, and research findings* (pp. 67-77). San Francisco, CA: Jossey-Bass Publishers.
- Phye, G. D. (1997). Classroom assessment: a multidimensional perspective. In G. D. Phye (Ed.), *Handbook of classroom assessment: learning, adjustment, and achievement* (pp. 33-51). San Diego, CA: Academic Press.
- Pryor, J., & Torrance, H. (2000). Questioning the three bears: the social construction of classroom assessment. In A. Filer (Ed.), *Assessment: social practice and social product* (pp. 110-128). London: RoutledgeFalmer.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13.

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32(2), 118-126.
- Schoenfeld, A. H. (2002). Making mathematics work for all children: issues of standards, testing, and equity. *Educational Researcher*, 31(1), 13-25.
- Shepard, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20(6), 2-16.
- Stefani, L. A. J. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23(4), 339-350.
- Stiggins, R. J. (1992). High quality classroom assessment: what does it really mean? *Educational Measurement: Issues and Practice*, 11(2), 35-39.
- Stiggins, R. J. (1997). *Student-centered classroom assessment* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Merrill.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26(3), 233-246.
- Swan, M. (2001). Dealing with misconceptions in mathematics. In P. Gates (Ed.), *Issues in mathematics teaching* (pp. 147-165). London: RoutledgeFalmer.
- Torrance, H. (1995). The role of assessment in educational reform. In H. Torrance (Ed.), *Evaluating authentic assessment: problems and possibilities in new approaches to assessment* (pp. 144-156). Buckingham: Open University Press.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- Walker, C., & Angelo, T. (1998). A collective effort Classroom Assessment Technique: promoting high performance in student teams. In T. Angelo (Ed.), *Classroom Assessment and Research: an update on uses, approaches, and research findings* (pp. 101-112). San Francisco, CA: Jossey-Bass Publishers.
- Watson, A. (2001). Making judgements about pupils' mathematics. In P. Gates (Ed.), *Issues in mathematics teaching* (pp. 217-231). London: RoutledgeFalmer.

- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: what's in it for schools?* London: RoutledgeFalmer.
- Wiliam, D. (1998). Enculturating learners into communities of practice: raising achievement through classroom assessment. Paper presented at European Conference on Educational Research, University of Ljubljana, Slovenia, 17-20 September 1998. Available online at: <http://www.kcl.ac.uk/depsta/education/publications/ECER98.pdf>
- Wiliam, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative from summative functions of assessment? *British Educational Research Journal*, 22(5), 537-548.
- Wood, R. (1990). The agenda for educational measurement. In T. Horton (Ed.), *Assessment debates* (pp. 48-56). London: Hodder & Stoughton.