The promise and potential of e-assessment for learning

Open Universiteit
www.ou.nl
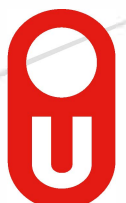
The promise and potential of e-assessment for learning

Eric Ras, Denise Whitelock, Marco Kalz

1 Introduction

Formative assessment has gained substantial ground in the last ten years, together with a number of considerable promises that have been made about its potential to promote student learning. The recent drive towards assessment for learning and assessment for 21st Century skills raises a set of new challenges for both teachers and students alike. These challenges are related, on the one hand, to progress monitoring which results in responsive teaching or support activities, on the other hand to the development and implementation of technologies that will allow (semi-) automated and personalised assessment systems. New data sources collected from such software will open new doors for formative assessment practices and related feedback types.

Hattie and Yates emphasise the role of feedback for learning. They underpin the importance of feedback in the learning process, because when feedback is provided in the right manner (i.e., level, timing, and quality), learners can profit substantially. Formative assessment has become more and more popular in the 1990s. Researchers and practitioners discussed its terminology, its components, its impact on learning as well as the influence of technology on formative assessment.

In general, technologies have shown a positive effect on learning achievements in education (e.g., an average effect size of 0.33 reported in (Tamin, Bernard, Borokhovski, Abrami, & Schmid, 2011). Hattie (2008) has conducted a review of meta analyses with regard to computer-assisted education over the past 30 years. The impact of computers on learning achievements was found positive with overall effect size of 0.37. These effects sizes were reported from all schooling and ability levels. The strongest effects were reported when computers supplemented traditional teaching, teachers were better trained to use them, student assumed to have control over their learning situation (pacing and mastering new material), working in pairs, and when computers were used to provide adaptive feedback.

Technology can play a double role in the process of supporting the development of effective formative assessment practices: Technology can be an enabler to deliver data as a basis to facilitate self-, peer- or teacher-driven assessment and feedback, or technology can create completely new assessment practices. Redecker & Johannessen (2013) describe the role of technology for assessment based on the SAMR-model ranging from two levels of enhancement (substitution and augmentation) to two levels of transformation (modification and redefinition).

Before we elaborate in this chapter on the promises of assessment for learning and related technologies, we summarise the history of formative assessment, discuss how the terminology has evolved and interrogate recent reviews of formative assessment in order to understand what we can expect from applying the key concepts of formative assessment to support learning. These insights help us to understand the importance of providing and applying feedback, the context in which this feedback takes place, as well as the requirements technologies need to address.

The following Section (2) introduces the core concept of assessment for learning which is integral to the notion of feedback (see Section 3). Both sections discuss the central formative assessment strategies employed in both schools and higher education. Section 4 identifies the promises that have been made in the last years about the potential benefits of assessment for learning and reviews the evidence which support and question these claims. Section 5 provides hands-on examples of e-assessment for learning. Section 6 concludes the chapter.

2 Formative assessment and assessment for learning – effects, strategies and principles

The early research reviews explained how assessment can inform instruction (e.g., Fuchs & Fuchs, 1986; Natriello, 1987; Crooks, 1988; Gipps, 1999) and there were mixed findings from these reviews. Fuchs et al reported positive effects while the others reported a negative effect on learning. Black and Wiliam (1998) then produced a review of 250 relevant studies and concluded that "despite the existence of some marginal and even negative results, …, significant gains can be achieved by many different routes, and initiatives here are not likely to fail through neglect of delicate and subtle features" (p. 61) and suggested an effect size between 0.4-0.7 for the effect on learning (Black & Wiliam, 1998). In addition, other studies have focused on the impact of feedback on learning (e.g., Kluger & DeNisi, 1996; Hattie & Timperley, 2007; or Shute, 2008). Wiliam (2011) referred to results from a review in a Master Thesis which reported similar results (0.4 standard deviations). This is similar to the effect sizes mentioned in Kluger and DeNisi (1996). Hattie and Timperley (2007) proposed an average effect size of 0.95 for those studies they labelled "Feedback".

The differences in the effect sizes were due to different educational levels, variability in the population investigated as well as different sensitivity of the measures to the effect of instruction (Wiliam, 2011b). As mentioned earlier, even negative effects have been reported (Wiliam, Lee, Harrison & Black, 2004). The different results have led to a lot of criticism and discussion about anappropriate methodology and confidence in effect sizes (see Baird, Hopfenbeck, Newton, Stobart, & Steen-Utheim, 2014, for a summary).

Despite the critics, Black, Harrison, Hodgen, Marshall, and Wiliam (2005) stated that

> formative assessment leads to improved student achievement in all classes, with all teachers on all occasions … Our claim is that formative assessment in general is an effective intervention, although we do not underestimate the difficulties in translating theory into practice. (p. 7)

Torrance (2012) also emphasised this point by saying that "developing and implementing formative assessment is generally regarded as a 'good thing'" (p. 329), despite the critics of other researchers. These studies were helpful to derive "standards" for assessment for learning and feedback, e.g., the *theory of formative assessment* (Wiliam & Black, 2009) or design principles for using formative e-assessment to support learner's self-regulation (Nicol & MacFarlane-Dick, 2006). McMillan, Venable & Varier (2013) emphasised that researchers

> need to clearly conceptualize and operationalize what formative assessment characteristics are used in their studies… otherwise it is difficult to get a more holistic perspective about the effect of different components of formative assessment (p. 6)

What is the origin of the term *formative assessment*? Scriven (1967) first used the term *formative evaluation* and defined it as "the evaluation of an ongoing and malleable educational program" (p. 2). Bloom (1969, p. 48) also used this distinction to explain "formative evaluation to provide feedback and correctives at each stage of the teaching-learning process". By formative evaluation Bloom (1969, p. 48) refers to the "evaluation by brief tests used by teachers and students as aids of the learning process".

In the course of all those studies and reviews, many definitions for formative assessment were constructed. Black and Wiliam (1998) defined formative assessment "as encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify teaching and learning activities in which they are engaged". The definitions that followed considered formative assessment as a process: "the process used by teachers and student to recognise and respond to student learning in order to enhance that learning, during the learning" (Cowie & Bell, 1966/1999, p. 3), "assessment carried out during the instructional process for the purpose of improving teaching or learning" (Shepard, 2005, p. 275), "frequent, interactive assessments of students' progress and understanding to identify learning needs and adjust teaching appropriately" (Looney, 2005, p. 21), "the formative use of assessment information" (Good, 2011, p. 5). Filsecker and Kerres (2012, p. 4) define formative assessment as a „series of informed and informing actions that change the current state of the reciprocal teaching-learning relationship toward

a more knowledgeable one". They emphasise that the commercial testing industry often uses the term formative assessment for assessments which are done regularly to inform educators about whether standards are mastered or not. In fact, these have nothing to do with day-to-day and individual-adjusted instruction. Although we will discuss Assessment FOR Learning and relevant strategies and principles in the following section of the chapter, it is important to recognise that Assessment OF Learning is a central process in education and training. The construction of valid and reliable tests is important because it has an impact on the results of assessment scores and the inferences we make from them (Wiliam, 2008). Hence, there are also interaction effects between Assessment of Learning and Assessment for Learning. However, we are going to examine Assessment for Learning more closely rather than Assessment of Learning.

The term "assessment for learning" was suggested by the Assessment Reform Group in the United Kingdom (Assessment Reform Group, 1999). The difference between formative assessment and assessment for learning was explained by Stiggins (2005). Formative assessment is meant to be frequent, about a continuous stream of evidence of student progress. Formative assessment links to standards, whereas in assessment for learning students partner with their teachers to continuously monitor their current level of achievements in relation to agreed-upon expectations. Students set their goals for what to learn next and thus manage their own progress. Assessment for learning is a journey of success, as stated by Stiggins (2005), and students become able to communicate evidence of being a learner to other students, teachers and parents. While formative assessment informs the teacher about student progress, assessment for learning informs the student about their own learning.

Getting back to the term formative assessment, the literal meaning of the word *formative* means *to form something*. Hence, formative assessment should shape instruction and improve learning. Baclk and Wiliam (2009) derived a definition for formative assessment from several practical assessment cases, which emphasises on the decisions that shape instruction

> An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have made in the absence of that evidence. (p. 9)

This definition shows clearly that learners and peers take decisions based on evidence. Instruction need to be understood as a combination of teaching and learning activities intended to increase capacities of an individual to act in valued ways.

Wiliam's definition brings the former definitions of formative assessment closer to assessment for learning as defined by Stiggins. Stiggins (2005) meanwhile also considers students to be data-based instructional decision makers that assessment for learning keeps them informed on where they are in relation to where they want to be. Students use evidence of the current progress to manage and adjust their own learning (Stiggins, Arter, Chappuis, & Chappuis, 2006).

Assessment of Learning can promote effective learning but it is not guaranteed that more testing will necessarily enhance learning (see Assessment Reform Group). The Assessment Reform Group in *Assessment for Learning: Beyond the Black Box* (1999) indicates that Assessment for Learning depends upon the following factors:
- The provision of effective feedback to pupils,
- The active involvement of pupils in their own learning,
- Adjusting teaching to take account of the results of assessment,
- A recognition of the profound influence assessment has on the pupils' motivation and self-esteem,
- The need for pupils to be able to assess themselves and understand how to improve.

The Assessment Reform Group has formulated ten principles for Assessment for Learning:
- Is part of effective planning,
- Focuses on how pupils learn,
- Is central to classroom practice,

- Is a key professional skill,
- Is sensitive and constructive,
- Fosters motivation,
- Promotes understanding of goals and criteria,
- Helps learners know how to improve,
- Develops the capacity for self (and peer) assessment,
- Recognises all educational achievement.

To connect well to the scope of the book the use of ICT in schools and the emergence of a "data culture" as an omnipresent phenomenon, we will use the concept of assessment for learning, since we envision that technologies in the future will enable a continuous stream of data, which is supposed to be continuously used to shape learning activities and instructional and teaching activities with the purpose to improve learning.

Several strategies for assessment for learning have been developed. They often orient themselves around the three questions formulated by Atkin, Black and Coffey (2001): Where am I going? Where am I now? How can I close the gap? Hattie and Timperley (2007) have developed a model of feedback along these three questions in order to address the gap between actual performance and goal attainment. The three questions address *feed up*, *feed back*, *feed forward* respectively. Feedback can have an impact on four levels: task, process, self-regulation and self level.

Black and Wiliam (2009, p. 8) came up with a similar framework along these three questions (Table 1) and five strategies (numbered in Table 1). They further distinguish the roles of teacher, peer, and learner.

Table 1 – Aspects of formative assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1. Clarifying learning intentions and criteria for success	2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3. Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4. Activating students as instructional resources for one Another	
Learner	Understanding learning intentions and criteria for success	5. Activating students as the owners of their own learning	

The authors highlight the role of peers as being important in the assessment process and in particular to support the teacher in their learning a continuous feedback stream. They emphasise as well that learners need to become owners of their own learning, which raises the issues of metacognition, interest, attribution, and self-assessment.

In a similar spirit, Chappuis (2009, p. 11) suggests seven strategies for assessment for learning:

Where Am I Going?

1. Provide students with a clear and understandable vision of the learning target
2. Use examples and models of strong and weak work

Where Am I Now?

3. Offer regular descriptive feedback
4. Teach students to self-assess and set goals

How Can I Close the Gap?

5. Design lessons to focus on one learning target or aspect of quality at a time
6. Teach students focused revision
7. Engage students in self-reflection, and let them keep track of and share their learning

While these strategies are proposed as general strategies Good (2011) points out that strategies interact tightly with the learning context and the learning content (see Fig. 1).
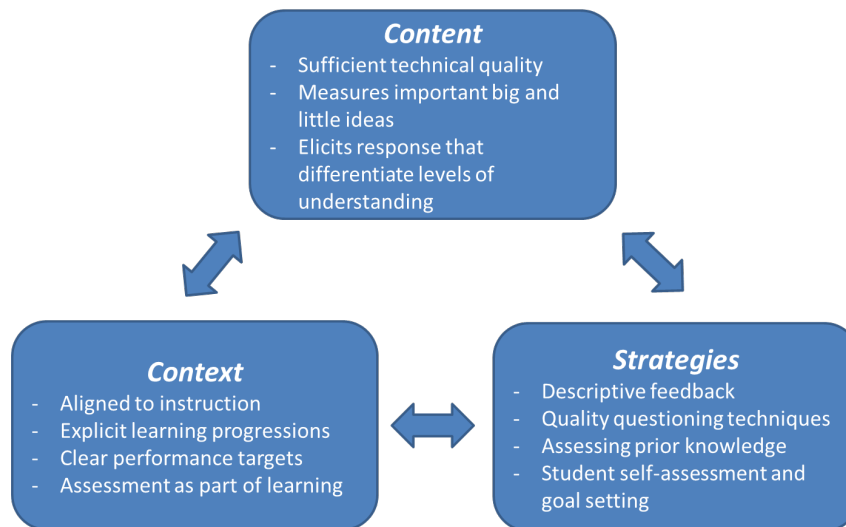

Figure 1. Components of a formative process

Note. Retrieved from Good (2011, p. 4), *Practical Assessment, Research & Evaluation* (PARE). Reprinted with permission.

The strategies mentioned in the model reflect the commonly known strategies used by other frameworks. Content refers to item quality and appropriate difficulty, reliability, lack of bias etc. and address different levels of understanding. Both teachers and learners need to comprehend the gap between the current state and the targeted goal as well as they need to receive information permitting them to take decisions regarding instruction and learning. The context therefore clearly states the learning goals and instructional targets. Information on learning progression for teachers and learners is essential and needs to be linked with short- and long-term learning objectives. In addition, learners need to understand that assessment is an essential part of learning and that they need to take responsibility for their learning.

Good's model is based on the *formative use of information* and shows that formative assessment goes far beyond the response to a particular assessment item. Learners and teachers should not rely soley on the results from assessments because this ignores the relationship between context, content, and strategies.

Even if more and more information becomes available in the educational context, it does not mean that it is used to adapt instruction or learning. The missing component for this is feedback being *exchanged* between teachers, learner and peers in the formative assessment process.

3 Feedback

Although formative assessment can be defined in a number of ways, its essential function is that it assists learning. Feedback originates from the field of engineering and information theory with the general assumption that information about the current system's state is used to change the future state. Hattie and Yates (2014, p. 66) consider feedback as empowering because it enables the learner to "move forward, plot, plan, adjust rethink and exercise self-regulation".

A core problem is that feedback is often only a transmission of information from the teacher to the student, without any measure to monitor whether the information is used and hence results in improved work or a change in learning. Ramaprasad (1983, p. 4) has very clearly stated: "The

information on the gap between the actual level and the reference level is feedback only when it is used to alter the gap". This is similar to the notion proposed by Whitelock (2010) which she terms as "Advice for Action". Also Boud and Molloy (2013a) emphasised that the focus should be on the use of feedback, not only the delivery of feedback. Boud and Molloy (2013a) define it

> as a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work. (p. 6)

As mentioned in the previous section Hattie and Timperley (2007) locate feedback on four different levels. The *task level* tackles the product or outcome of a learning activity and feedback is related to whether it is correct or not. The *process level* refers to the process of creating the product or accomplishing a task. Feedback can refer to processing of information in a learning task, following a different solving strategy, or summarising a text differently. Feedback on the *self-regulation level* helps the student to judge his self-regulation and confidence skills with regard to a specific task. Feedback on the *self level* is personal and normally unrelated to the performance on the task (e.g., praise).

Even with the help of digital resources and technology, Boud and Molloy (2013a) point out that relying fully on teachers to keep the feedback loop alive is simply not sustainable. They suggest giving the learners more agency to ameliorate the dependence on teachers or the teaching system. They also proposed a list of features for feedback with long term effects, i.e., leading to transferrable capabilities which can be used beyond graduation. Boud and Molloy (2013a) categorise their features around three elements a) learners self-regulating their learning, b) curriculum, c) learning milieu:

a) Nicol stated that

> when students receive feedback from teachers they must engage in self-assessment if they are to use that information to improve academic performance: that is, they must decode the feedback message, internalise it and use it to make judgements about and modify their own work. (p. 339)

In order to support self-regulated learning, Nicol and McFarlane (2006, p. 203) listed seven principles of good feedback practice based on the model originally published by Butler and Winne (1995):

1. clarify what good performance is (goals, criteria, expected standards);
2. facilitate the development of self-assessment (reflection) in learning;
3. deliver high quality information to students about their learning;
4. encourage teacher and peer dialogue around learning;
5. encourage positive motivational beliefs and self-esteem;
6. provide opportunities to close the gap between current and desired performance;
7. provide information to teachers that can be used to help shape teaching.

Boud and Molloy (2013b) expect the higher education students moving from a sole recipient of information provided by others to an active learner eliciting knowledge for improvement. This means that teachers need to understand that feedback is oriented towards self-regulation and that students become able to make judgements and to act upon them.

b) With regard to the curriculum, Boud and Molloy expect the learners to interact in a learning environment where they judge their work and compare it to external appraisals within the domain of their studies. "Feedback becomes a key curriculum space for communicating, for knowing, for judging and for acting" (2013a). This means that learners need to understand the purpose of feedback and learning outcomes as support in the development of judgements and in the collaboration with peers. Learners acquire feedback seeking skills (e.g., identification of criteria, formulating comments on other's work) and enrolling in tasks which increase their complexity step by step.

c) The implementation of the curriculum needs to provide a learning milieu where a dialogue between learners, teachers, peers and systems to help learners is fostered (see also McArthur, Huxham, 2013; Merry, Price, Carless, & Taras, 2013 on feedback as dialogue). It refers to the climate of cooperation between students as well as the issue of building a relationship of trust. Teachers become designers and sustainers of the learning milieu. They create the conditions in which students can operate with

agency. Nevertheless, Boud and Molloy (2013b) emphasise that it does not mean that feedback provided by the teacher becomes redundant. Students still need information about where they stand compared to standards and information which enable them to refine their own judgements. The focus of feedback needs to shift towards better quality information about students learning for teachers and students need "to better exercise their skills in eliciting the kinds of information they need".

The verdict on feedback however, as mentioned earlier, is not entirely positive. Kluger and DeNisi (1996) offered a critical review and meta-analysis of the feedback literature, stating that a positive effect of feedback on subsequent performance is far from universal, even though it may appear to be represented this way in the literature. Whilst Kluger and DeNisi's paper was written almost 20 years ago, it is important that we still heed its warning not to assume feedback will automatically be beneficial. However, they do say that a *feedback intervention* will be beneficial if it supports learning and helps learners to set goals (Kluger & DeNisi, 1996). This is further supported by Hattie and Timperley's (2007, p. 89) conclusion that "When goals have appropriate challenge and teachers and students are committed to these goals, a clearer understanding of the criteria for success is likely to be shared". But which types of feedback are beneficial?

Chi, Siler, Jeong, Yamauchi, and Hausmann (2001) also assert that 'suggestive feedback' is helpful to learners, by highlighting an area that may be in need of work and so encouraging students to reflect on their work without directly giving the answer. Quintana, Zhang and Krajcik's (2005) review of software feedback reveals that hints have been given as responsive prompts, at assumed points of need during particular activities. These types of hints tend to be of short duration, in response to a potential error, omission or misunderstanding. Hattie and Timperley (2007) also offer a review of this research field and conclude that provision of *cues* can offer guidance to students for ongoing work:

> Such cues sensitize students to the competence or strategy information in a task or situation. Ideally, it moves from the task to the processes or understandings necessary to learn the task to regulation about continuing beyond the task to more challenging tasks and goals. (p. 102).

Whitelock (2010) has argued that feedback is rather restrictive in nature when formative assessment's focus is that of "Assessment for Learning". She suggests that what is required in this context is a concept known as "Advice for Action". This approach does not restrict itself to giving advice after a task has been completed but can also embrace hints given before an assessment task is taken up.

Shute (2008) derived guidelines for formative assessment and feedback to enhance learning. These guidelines are (summarised from Shute, 2008, Table 2, p. 177):

- feedback should be elaborated and focus on the task and not on the learner,
- It should be delivered after the attempt to solve a problem in manageable units to avoid cognitive overload,
- Feedback is clear and specific and links to performance and goals,
- Feedback should reduce uncertainty with regard to how well the learner performed on a task,
- We should use feedback to promote a "learning" goal orientation.

According to Shute, we must avoid providing feedback which compares learners or feedback providing overall grades. Praise should be used sparingly or not at all and we should avoid feedback that discourages the learners or threatens the learners' self-esteem. Feedback should not be provided when the learners are actively engaged. We should apply additional modes other than text to deliver feedback. Prompts and cues should be used instead of progressive hints which terminate with the correct answer.

With regard to timing, Shute suggests to use delayed feedback for simple tasks and immediate feedback for complex tasks. Furthermore, immediate feedback supports the retention of conceptual and procedural knowledge. Learner characteristics need to be considered as well (see also Narciss & Huth, 2004). Facilitative feedback is more useful for high-achieving learners when shown delayed, whereas directive (corrective) feedback should be delivered immediately to low achieving learners. In general, low performers profit from scaffolding and using feedback of the type correct response or

elaboration. For high performers only verification feedback might be enough. Learners who tend to a low learning orientation need specific and goal-directed feedback.

Shute (2008) concludes that information about learner characteristics and about desired outcomes can be used to develop *adaptive formative feedback* personalised to the learner and adapted to the nature of the task. Feedback can be generated on the fly based on a formative feedback model, which incorporates learning characteristics as well as the instructional context.

In their work with 10-14 year old school students, Narciss et al. reported a randomised control trial on the automated provision of 'hints' within short math tasks (Narciss, 2013; Narciss et al., 2014). Hints were provided after errors had been made in a task, but prior to a further attempt at the same task. The hints were therefore pre-emptive, to support future performance and learning, but were also a direct response to an error. Hints were designed to offer either conceptual or procedural guidance. In doing this work Narciss recognised that there is little research, theoretical or empirical, on 'automatic feedback adaptation', which is similar to our interpretation of the existing literature. Given the nature of the tasks tested within Narciss' studies, being in the math domain and specifically working with fractions, students' responses were relatively easy to identify as correct or incorrect. As Narciss acknowledges, this is not the case within less-structured tasks, such as essay writing, and so the nature of feedback needed is significantly different.

In her review on feedback in assessment, Evans (2013) synthesised the principles of effective feedback practice and feed forward from hundreds of studies in higher education. The principles range from those addressing the needs of students, explaining the learning environment, to institutional changes and training of teaching staff. What is interesting is that Evans has identified *e-assessment feedback* as one of the three core themes, besides self-feedback and peer-feedback, which increasingly show up in scientific literature more recently.

For a deeper discussion of the link of formative assessment to the different theories (i.e., behaviourist, cognitive, constructivist, social constructivist, socio-cultural) and its implications, see the recent review on assessment and learning (Baird et al., 2014). Despite all efforts to define formative assessment and investigate the effect on learning, Baird et al. also summarised the critiques made about the definitions proposed, the effect sizes found, the underrepresentation of measurement principles, etc. by relying mostly on Bennett's critical review on formative assessment (Bennett, 2011).

To close this section, we refer to Boud and Molloy (2013b) who stated that format assessment should shift "feedback from a notion of telling followed by identifiable utilisation to one of seeking followed by judgement and identifiable utilisation". After having introduced assessment for learning and feedback, we will elaborate more on recent studies that report empirical evidence of *e-assessment on learning*.

4 E-assessment for learning – State of the art

Pachler, Daly, Mor, and Mellar (2010) define formative *e-assessment* as

> the use of ICT to support the iterative process of gathering and analysing information about student learning by teachers as well as learners and of evaluating it in relation to prior achievement and attainment of intended, as well as unintended learning outcomes. (p. 716)

Bull and McKenna (2004, p. 12) stated that computer-assisted assessment "offers as sort of bridge between summative and formative assessment". Hattie and Timperley (2007) report an effect size of 0.53 for *computer-assisted instructional feedback*. Stödberg (2011) reports that for the period 2004-2009, out of 76 articles, 29 were concerned with formative assessment and 17 articles consider both summative and formative e-assessment. His review revealed that most e-assessment tasks were built on closed questions. In addition, portfolios and discussions were frequently used in formative e-assessment. An interesting finding was that only a few studies were found which support automation throughout the assessment process, and that peer assessment is a common element to represent the social dimension of e-assessment.

In their report on online formative assessment Gikandi, Morrow, and Davis (2011) selected 18 key studies, most of them case studies. They suggest that different techniques (peer-, self-assessment, etc.) are operationalised through a systematic usage of different online tools (e.g., asynchronous online discussion tools, self-test quiz tools, e-portfolios). The authors conclude that further research is necessary to develop strategies about "which tools support the optimum level of meaningful interactions and other valuable experiences for online learners within various disciplines and particular subject areas", and which configurations for these tools effectively facilitate the desirable formative processes.

Whitelock, Gilbert, and Gale (2011) report on the desktop research commissioned by the Higher Education Academy which set out to consult with the academic community about which references on assessment and feedback with technology enhancement were most useful to practitioners. The key message from these sources was that case studies rather than controlled experiments were influencing practice. The authors found that most of the recommended literature focused on the goals that technology enhancement can enable assessment and feedback to meet, and on how assessment and feedback can be designed to make best use of the technology.

Some of the most successful case studies were about learning designs that use technology enhancement successfully. An example of how a successful learning design can be enhanced by technology is provided in Crouch and Mazur (2001). Their paper describes the results of ten years' experience of improved student results (compared with traditional instruction and therefore in evidence category 1b) using a method they call *Peer Instruction*:

> A class taught with PI [Peer Instruction] is divided into a series of short presentations, each focused on a central point and followed by a related conceptual question [MCQ example given]. Students are given one or two minutes to formulate individual answers and report their answers [using a poll] to the instructor. Students then discuss their answers with others sitting around them; the instructor urges students to try and convince each other of the correctness of their own answer by explaining the underlying reasoning. Finally, the instructor […] polls students for their answers again (which may have changed based on the discussion), explains the answer and moves on to the next topic. (p. 970)

The authors found that the "vast majority" of students who changed their vote after the peer discussion moved from an incorrect answer to the correct answer. Draper (2009) discusses how this technique can be used with an electronic voting system, a technology used to display the question, capture the student responses, and display the votes for each option as a graph.

Technology enhancement is not just applied to MCQs, however. Jordan and Mitchell (2009) provide evidence for moving beyond the MCQ and using open questions with technology enhancement. They suggest that open questions are suitable for computerised delivery and feedback "if correct answers can be given in short phrases or simple sentences and the difference between correct and incorrect answers is clear-cut" (p. 382). Whitelock and Watt (2008) illustrate this effect using the Open University's *Open Comment* system.

Ashton, Beevers, Korabinski, and Youngson (2006) provide evidence that technology-enhanced methods can be used to mirror tutor marking practices in mathematical examinations. They explain how software was developed and how some questions were redesigned to allow partial credits to be awarded and mathematical expressions to be entered by students in automated exams.

Boyle and Hutchinson (2009) address the issue of whether or not sophisticated tasks can be assessed using technology enhancement. They suggest that

> e-assessment will become an important and widely-used feature of education systems in the near future. Further, the types of questions and tasks used in near-future e-assessment may well be quite different from questions and tasks used in on-paper assessment, and in early implementations of computerised assessment. (p. 306)

An example of innovative items types can be found in Ras, Krkovic, Greiff, Tobias, and Maquil, (2014). They developed items based on linear equations to assess complex collaborative problem solving skills. The items were implemented on a tangible user interface, where learners interact with

physical objects on an interactive surface. The table tracks each interaction and provides feedback based on the physical manipulation of several environmental parameters.

After having summarised the general patterns of e-assessment development and after summarising some recent efforts we should always keep in mind, that assessment technology is and can never be neutral and that there are always effects on the psychological and social level. Daly, Pachler, Mor, and Mellar (2010) argue that "effective e-assessment needs to take account of the human-centric, social dimension as well as technological, data-gathering and management perspectives" (p. 620).

5. E-Assessment for learning – Promises, potentials and future development

Having now introduced the foundation of assessment for learning and feedback and examples of current research, the question arises which role technology can play in innovation in the assessment domain, with a special focus on assessment for learning. Chudowsky and Pellegrino (2003) have stated that

> by enriching assessment situations through the use of multimedia, interactivity, and control over the stimulus display, it is possible to assess a much wider array of constructs than was previously possible. New capabilities afforded by technology include directly assessing problem-solving skills, making visible sequences of actions taken by learners in solving problems, and modeling complex reasoning tasks. (p. 79)

Pellegrino, Chudowsky, and Glaser (2001) have earlier formulated their vision with regard to future technological development and the availability of data from assessment

> One can imagine a future in which the audit function of large-scale external assessments would be significantly reduced or even rendered unnecessary because the information needed to assess students, at the levels of description appropriate for various assessment purposes, could be derived from the data generated by students in and out of their classrooms. Technology could offer ways of creating over time a stream of data about how students think and reason while engaged in important learning activities. Information for assessment purposes could be extracted from this stream and used to serve both classroom and external assessment needs, including providing individual feedback to students for reflection about their states of knowledge and understanding. (p. 284)

The development of technology for assessment purposes is of course embedded into the general discussion about the role of assessment. Bennett (1998) describes three generations of e-assessment: First Generation tests using designs based closely on existing paper-based tests, Next-Generation tests which use new formats including multimedia, constructed response, automatic item generation and automatic scoring, and Generation ''R'' (Reinvention) tests which use complex simulations and intelligent tutors (Bennett, 1998). In a recent paper Redeker and Johannessen (2013) have drafted a framework for the development of technology-enhanced assessment innovations. According to the authors, e-assessment has developed from a focus on (adaptive) testing, over a process of re-inventing the testing paradigm with technology which moves towards an embedded assessment. The authors expect that in the embedded assessment paradigm, data are constantly produced via learning analytics rather than through moments of (formal) testing. The authors conclude that new technological developments are leading the way towards a new assessment paradigm but that pedagogy is lagging behind these developments. While we are sceptical about this deterministic stance, we agree with the authors that the innovation process for e-assessment can be exploratory and technology driven, or driven by new educational demands and learning practices. In essence we support this view with the extracts of six components of innovative e-assessment for learning approaches from the literature and state of the art developments, which include the following:

1. *Agency change*: While many e-assessment technologies are still rooted in an old testing paradigm triggered by the institution or the teacher, new approaches need to strife for an agency change towards the learners as the trigger of feedback and assessment processes. This component in particular addresses the requirement to support the learner's self-regulation. Following the recommendations by strategies and principles introduced earlier, agency change means also that

a learner should seek (pull) for information followed by judgement and identifiable utilisation instead of a assessment process of telling (push) the learner about the process and goal achievements.

2. *Flexible timing*: Future assessment and feedback needs to be available when needed by the learner and must avoid disturbing the learner in the learning process. Furthermore, timing depends on the learner's characteristics (e.g., performance level, goal orientation level) and the complexity of the task.

3. *Automation*: To avoid an overload of teachers and learners automation is important. Automation can happen at design time of the assessment, during run-time (i.e., solving the test item including feedback mechanism, during scoring, or even after the feedback has been provided. Scoring is meant to be the evaluation of the student's answer to an assessment item whereas the last category of automation refers to identifying the utilisation of feedback.

4. *Adaptivity/Adaptability*: Assessment and feedback needs to be adaptive towards the individual and his state of knowledge and other preferences. Adaptability means that the personalisation are controlled and steered by the user (i.e., user-driven). Adaptivity means that the system controls the personalisation (i.e., system-driven).

5. *Data triangulation*: Scoring and rich feedback need to combine data from different sources.

6. *Continuity and dialogue*: Feedback and assessment needs to be a continuous process and not restricted to ongoing courses or the schedule of the study year. A continuous dialogue between teachers, learner, peers and systems is essential.

In the following part we introduce three case-studies which represent innovative future directions for e-assessment and assessment for learning. The following table depicts the cases and the aspects that these cases cover in terms of assessment innovation.

Table 2 – E-assessment cases and their innovation perspective

	Agency change	Flexible timing	Automation	Adaptivity/ Adaptability	Data triangulation	Continuity and dialogue
Case 1: SafeSea	++	+++	+++	++	++	++
Case 2: Ubiquitous Assessment and Feedback	++	+++	+	0	++	++
Case 3: Automatic item generation	0	0	+++	++	0	+

(0 no support, +low, ++medium, +++strong support)

5.1 Case study – SAFeSEA

The SAFeSEA project (Supportive Automated Feedback for Short Essay Answers) focuses on the use of a Natural Language Analytics engine to provide direct feedback to students when preparing an essay for summative assessment. The challenge was to provide meaningful feedback to the students themselves so that they can self-correct rather than providing a recommender system which elicits a tutor intervention with the student (Arnold & Pistilli, 2012). Open University (OU) students used a computerised system (OpenEssayist) that provided automated feedback on draft essays, developed as part of the SAFeSEA project, to seek support with their essay writing skills.

OpenEssayist is a real-time learning analytics tool, which operates through the combination of a linguistic analysis engine, which processes the text in the essay, and a Web application that uses the output of the linguistic analysis engine to generate the feedback. The reason OpenEssayist was built

was because many students come to the OU to return to study after some time spent in the workforce, and so it is common that a significant period of time has passed since their last experience of writing academic essays. It is therefore not surprising that many find this task difficult, and without adequate support may decide to quit their course (Simpson, 2003). This is one crucial reason why a system that can intervene and offer support between students' draft and final submitted essays might be so valuable for students and tutors alike. In creating a system that can go some way to meeting these needs, a number of preliminary studies were made (Alden et al., 2014; Alden, Whitelock, Richardson, Field, & Pulman, 2014; Field et al., 2014; Richardson et al., 2014).

The final system was then developed to process open-text essays and which offers feedback to through key phrase extraction and extractive summarisation. Key phrase extraction identifies which individual words or short phrases are the most suggestive of an essay's content, while extractive summarisation essentially identifies whole key sentences. This operates under the assumption that the quality and position of key phrases and key sentences within an essay (i.e., relative to the position of its structural components) might give an idea of how complete and well-structured the essay is, and therefore provide a basis for building suitable models of feedback. Prior to this stage, each essay is automatically pre-processed using modules from the Natural Language Processing Toolkit. These modules include several tokenisers, a lemmatiser, a part-of-speech tagger, and a list of stop words.

Based on these extractive processes, the system then presents users with feedback on their writing in a number of different ways, including identification of the essay's most prominent words, with graphical illustrations of their use across the essay; identification of the essay's most representative sentences, with hints encouraging the user to reflect on whether these key sentences express, in their view, the central ideas of the essay; and graphical illustrations of the essay's internal structure. Users can move between the different representations, drawing on the features they find most informative and helpful in drafting their work.

OpenEssayist is unique in being a content-free tool that has been developed to offer automated feedback on students' draft essays, rather than an assessment on their finished work. OpenEssayist is a system that offers opportunities for students to engage with and reflect on their work, in any subject domain, and to improve their work through understanding of the requirements of academic essay writing. In trial use of the system in a genuine Open University course, we found that students made use of it to varying degrees, which is perhaps likely with any study resource. Those who took the time to explore system affordances and what they could be used for however tended to report more positively on its perceived value. From our analysis we were also able to conclude that a significant positive correlation exists in this sample of students between grades on essay and the number of drafts submitted. We also found that students who had access to OpenEssayist achieved significantly higher grades for this course than the previous year of students, who had no such access. We could speculate as to what this may mean for this set of students, or more widely, but it seems clear that use of a system such as OpenEssayist has many potential advantages to students and tutors, which will benefit from further research and exploration.

Moving forward, as OpenEssayist is designed to offer feedback to students during the drafting process, this has considerable implications for supporting students to improve their work, and also supporting students to *believe* that they can improve their academic work. This is no small feat for learners who may often feel isolated and stretched trying to squeeze study around other commitments and demands on their time.

5.2 Case study – Ubiquitous assessment and feedback

Ubiquitous and mobile technologies have the potential to be important driver for change in terms of assessment and feedback innovation. While learners are already using mobile devices and sensors to track their performance in sports, leisure activities or activity patterns for health or pro-environmental behaviour (also known as 'quantified self', see Swan, 2012) the use of mobile devices for learning, assessment and feedback is heavily underexploited. We have recently explored the use of ubiquitous technologies for feedback and assessment in different ways.

In a recent study we have employed mobile notifications to trigger meta-learning and meta-cognitive development (Tabuenca, Kalz, Ternier, & Specht, 2014). In this study we have assessed in how far notifications on mobile devices are suited to trigger reflective processes and meta-cognitive learning. The study has shown that the most promising way to initiate reflection about learning is to employ notifications in combination with reflection triggers ex post or as reflection-on-action (Schön, 1983). Regarding the questions how notifications are triggered, the study has shown that fully automated notifications resulted in disruptions of the learning experience and were not perceived as a productive process by participants of the study. Notifications triggered by the learners were perceived as a positive direction for future research and development. This is in line with the idea of agency change and the goal to give more control to the learner in the assessment and feedback process.

In another on-going study we are analysing how the active registration of time in combination with automated learning analytics statistics has an impact on meta-cognitive development and time-planning. Participants use their mobile devices to register the time they invest in learning activities of a specific course in a distance-teaching context. Different treatments for the learning analytics visualisations are tested in the study, and their impact on meta-cognitive skills and time planning is assessed.

In the context of lifelong learning one of the challenges is to integrate the scattered learning episodes into a continuous learning process (Kalz, 2015). For this purpose, we are exploring how learners can actively use sensors and mobile devices to model and adapt the feedback process depending on timing and location. The lifelong learning hub is an ecosystem designed to lead the lifelong learner towards a self-regulated process, to foster awareness on learning goals and learning moments and to facilitate the user to keep track of learning time with a frictionless interface (Tabuenca, Kalz, Specht, 2014a). This concept of the lifelong learning hub is currently extended to include different sensors (NFC, RFID) and interaction options to register data and trigger activities (Tabuenca, Kalz, Specht, 2014b).

This research direction is actively putting the learner into control of the data-collection process and the triggering of feedback. While at the moment the feedback produced is only at the level of the learning process and only fully automated, different other feedback layers could be added in the future and data from other systems could be integrated into the feedback process to provide a triangulation of data used for the feedback as well. It will be a balancing act to include feedback sent by teachers into the feedback process without breaking the continuity aspect.

5.3 Case study – Automated item generation

A lot of items are acquired not only for computerised adaptive testing in daily classroom settings, teachers are faced with the challenge to create a huge amount of items as well in order to provide a continuous assessment for learning. Other reasons such as reducing bias or providing a variety of test item types produce high costs for educational institutions.

In the context of e-assessment in particular, teachers need technical skills as well as pedagogical and psychometric skills to produce high quality items. That is why automatic item generation (AIG) becomes more and more popular in e-assessment: test items are generated automatically from templates, which are typically the item we know from assessment such as MCQ, cloze or match items. In these templates, the differences in stem, options and auxiliary information amongst items are specified using variables (Gierl & Haladyna, 2013; Gierl & Lai, 2013).

Different approaches exist to generate items. We classify them as either top-down or bottom-up approaches. Top down approaches follow an engineering process (Luecht, 2013). First the constructs to be assessed are defined. They include proficiency levels and also evidence models, which define how to reach the different levels. Task models detail further the skills and knowledge assets on the different levels. Item templates are connected to the task models and can be instantiated to produce items. Typically, items are stored in a so-called item repository to make them available to a larger community. The bottom-up approach is starting from an existing resource to generate items. An example can be a piece of learning material which a learner is currently using (Karamanis, Ha, & Mitkov, 2006). Another promising approach is to derive item from domain models. Foulonneau (2011), Linnebank, Liem, and Bredeweg (2010), Liu (2009), and Papasopoulos, Demetriadis, and

Stamelos (2009), for example, create items from semantic models in domains of medicine, history, environment, and computer science.

In order to assess publicly available semantic models from the Web, Foulonneau and Ras (2013) conducted a case study to investigate their usefulness for item generation. For this bottom-up approach they looked for semantic patterns in different ontologies (e.g., Infobox, DBpedia, FOAF, YAGO) and mapped them to the different knowledge levels (e.g., factual, conceptual). The analysis revealed that the model from the semantic Web can be used to assess factual knowledge and in some cases conceptual knowledge, because the related resources are mostly of descriptive nature. Nevertheless, the much richer domain models defined by domain experts have the potential to assess higher level knowledge since they often describe process, cause effect relationships etc. AIG is still at its beginning and the evolvement of such semantic models will certainly provide additional opportunities to create more sophisticated items. Foulonneau et al. have recently summarised all prospects of using open educational resources and the semantic Web for AIG (Foulonneau & Ras, 2014). In addition, first efforts have been made to generate feedback elements from textual resources and work has started to link item quality with resources metrics (e.g., link item difficulty with text complexity).

With regard to the six components for innovative e-assessment for learning approaches, the generated items can support the teacher in keeping up continuous assessment activities, which are also important for self-regulation. AIG approaches certainly contribute strongly to the automation aspect since they support teachers in effort-intensive item design tasks. Templates can be reused and the teacher can concentrate on adapting existing variable definitions and related patterns in the models or simply search for items in an item repository. Adaptive assessment systems require a critical mass of items, and that is why AIG indirectly supports adaptive systems, which for example aim at delivering adapted tests to the learner. If the item delivery is done through a professional assessment platform (e.g., TAO[i]) a stream of assessment data can be stored, analysed and made available for the teachers, learners, and other systems (e.g., learning managements systems, intelligent tutoring systems) which is the basis for a dialogue.

6. Conclusion and limitations

In this chapter we have introduced the current state-of-the art of e-assessment for learning. While we have reviewed large amounts of available literature, we have not discussed the potential interaction effects between assessment for learning and assessment of learning. In the future, an integrated assessment concept for educational institutions needs to take into account ways to combine both types of assessment and to realise a continuous process that is mainly steered by the learner. In the context of lifelong learning it will be also a challenge to overcome a strict connection of feedback to educational institutions and models, and technologies will need to be developed that allow a cross-institutional perspective for the learner.

The previously mentioned case studies are just single innovative approaches of assessment for learning. Many other approaches which address one or more of the previously listed components of innovative e-assessment approaches exist (see for example also (Pachler et al., 2010) for the use of audiofiles, wiki, mobile devices, open mentor, string comparison in e-assessment). What is important to understand is the move to adapt technological tools to meet pedagogical imperatives.

As mentioned earlier, technology can play a double role in developing effective assessment for learning practices: Technology can be an enabler to deliver data as a basis to facilitate self-, peer- or teacher-driven assessment and feedback or technology can create completely new assessment practices or allow the assessment of skills for which no assessment means have existed before.

Pachler et al. (2010) define the domain of formative e-assessment as extremely complex, because it is embedded into the process of teaching and learning and because technology "reshuffles the context of teacher-student interaction". As can be seen in Table 2, technologies are not able to directly address timing or data triangulation. It is the combination of technology, teachers, and learners in an appropriate learning milieu which leads to the achievement of such requirements. This chapter has

summarised the main assessment strategies and principles and they need to be considered when e-assessment technologies are developed and deployed in a learning context. "It is the learners and teachers as human actors who ultimately determine the formative effects of engaging with technologies, but technologies can shape the potential for this to happen (Pachler et al., 2010, p. 721)"

Technology has the potential to build on a number of research findings about feedback and self-reflection. Automated systems can provide Advice for Action and also review what steps have been taken to improve an assignment as was the case of OpenEssayist which was able to correct and save a number of draft essays. Automatic feedback systems can start to produce meaningful dialogues between students and teachers. What is perhaps more important is an ability to envision the pedagogical principles that will promote e-assessment for learning and then to produce user-centred designs that students can utilise throughout their self-regulated learning journeys.

References

Alden, B., Van Labeke, N., Field, D., Pulman, S., Richardson, J. T. E., & Whitelock, D. (2014). Using student experience as a model for designing an automatic feedback system for short essays. *International Journal of e-Assessment, 4*(1), article no. 68.

Alden, B., Whitelock, D., Richardson, J. T. E., Field, D., & Pulman, S. (2014). Functional, frustrating and full of potential: learners' experiences of a prototype for automated essay feedback. In: Kalz, M. & Ras, E. (eds). *Computer Assisted Assessment: Research into e-Assessment.* Communications in Computer and Information Science (439). Cham, Switzerland: Springer, 40-52. http://dx.doi.org/10.1007/978-3-319-08657-6_4

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. Paper presented at the *2nd International Conference on Learning Analytics and Knowledge*, April 29th – May 2nd, Vancouver, BC, Canada. ACM 978-1-4503-1111-3/12/04.

Ashton, H. S., Beevers, C. E., Korabinski, A. A., & Youngson, M. A. (2006). Incorporating partial credit in computer-aided assessment of Mathematics in secondary education. *British Journal of Educational Technology, 37*(1), 93-119. http://dx.doi.org/10.1111/j.1467-8535.2005.00512.x

Assessment Reform Group (1999). *Beyond the Black Box*. Retrieved from http://assessmentreformgroup.files.wordpress.com/2012/01/beyond_blackbox.pdf

Atkin, J. M., Black, P., & Coffey, J. (2001). *Classroom assessment and the national science standards*. Washington, DC: National Academies Press.

Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review assessment and learning*. Oxford: Oxford University, Centre for Educational Assessment.

Bennett, R. E. (1998). *Reinventing assessment: Speculations on the future of large-scale educational testing*. Princeton, NJ: Educational Testing Service Policy Information Center.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Black, P., Harrison, C., Hogden, J., Marshall, B., & Wiliam, D. (2005). Dissemination and evaluation: A response to Smith and Gorard. *Research Intelligence, 93*(7).

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5-31. http://dx.doi.org/10.1007/s11092-008-9068-5

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment and Evaluation: Principles, Policy & Practice, 5*(1), 7-73. http://dx.doi.org/10.1080/0969595980050102

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In H. G. Richey & W. D. Tyler (Eds.), *Educational evaluation: new roles, new means: the 63rd yearbook of the National Society for the Study of Education (part II), 69*(2), pp. 26-50). Chicago, IL: University of Chicago Press

Boud, D., & Molloy, E. (2013a). *Feedback in higher and professional education - Understanding it and doing it well*. New York, NY: Routledge.

Boud, D., & Molloy, E. (2013b). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698-712. http://dx.doi.org/10.1080/02602938.2012.691462

Boyle, A., & Hutchison, D. (2009). Sophisticated Tasks in E-Assessment: What Are They and What Are Their Benefits? *Assessment & Evaluation in Higher Education, 34*(3), 305-319. http://dx.doi.org/10.1080/02602930801956034

Bull, J., & McKenna, C. (2004). *Blueprint for computer-assisted assessment*. London: Routledge Falmer.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning - A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281. http://dx.doi.org/10.3102/00346543065003245

Chappuis, J. (2009). *Seven Strategies of Assessment for Learning - A Study Guide*. Portland: Pearson.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471-533. http://dx.doi.org/10.1207/s15516709cog2504_1

Chudowsky, N., & Pellegrino, J. W. (2003). Large-Scale assessments that support learning: What will it take? *Theory Into Practice, 42*(1), 75-83.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. Assessment in Education: Principles, Policy & Practice, 6(1), 101-116. http://dx.doi.org/10.1080/09695949993026 Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481. http://dx.doi.org/10.3102/00346543058004438

Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69*, 970-977. http://dx.doi.org/10.1119/1.1374249

Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring Formative E-Assessment: Using Case Stories and Design Patterns. *Assessment & Evaluation in Higher Education, 35*(5), 619-636.

Draper, S. (2009). Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology, 40*(2), 285-293. http://dx.doi.org/10.1080/02602931003650052

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research, 83*(1), 70-120. http://dx.doi.org/10.3102/0034654312474350

Field, D, Richardson, J. T. E., Pulman, S., Van Labeke, N., & Whitelock, D. (2014). An exploration of the features of graded student essays using domain- independent natural language techniques. International Journal of e-Assessment, 1(1).

Filsecker, M., & Kerres, M. (2012). Repositioning Formative Assessment from an Educational Assessment Perspective: A Response to Dunn & Mulvenon (2009) *Practical Assessment Research & Evaluation, 17*(16).

Foulonneau, M., & Ras, E. (2013). Using educational domain models for automatic item generation beyond factual knowledge assessment. Paper presented at the 8th European Conference on Technology Enhanced Learning (ECTEL 2013), Paphos, Cyprus.

Foulonneau, M., & Ras, E. (2014). Automatic Item Generation - New prospectives using open educational resources and the semantic web. *International Journal of e-Assessment (IJEA), 1*.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation. *Exceptional Children, 53*(3), 199-208.

Gierl, M. J., & Haladyna, T. M. (2013). Automatic Item Generation - An introduction. In M. J. Gierl & T. M. Haladyna (Eds.), *Automatic Item Generation*. New York, NY: Routledge.

Gierl, M. J., & Lai, H. (2013). Using weak and strong theory to create item models for automatic item generation. In M. J. Gierl & T. M. Haladyna (Eds.), *Automatic Item Generation*. New York, NY: Routledge.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333-2351. http://dx.doi.org/10.1016/j.compedu.2011.06.004

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, 24*, 355–392. http://dx.doi.org/10.3102/0091732X024001355

Good, R. (2011). Formative Use of Assessment Information: It's a Process, So Let's Say What We Mean. *Practical Assessment Research & Evaluation, 16*(3), 1-6.

Hattie, J. (2008). *Visible Learning*. New York, NY: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. http://dx.doi.org/10.3102/003465430298487

Hattie, J., & Yates, G. (2014). *Visible learning and the science of how we learn*. New York, NY: Routledge.

Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology, 40*(2), 371-385. http://dx.doi.org/10.1111/j.1467-8535.2008.00928.x

Kalz, M. (2015). Lifelong learning and its support with new technologies. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Oxford, United Kingdom: Pergamon.

Karamanis, N., Ha, L. A., & Mitkov, R. (2006). Generating multiple-choice test items from medical text: A pilot study. Paper presented at the Fourth International Conference Natural Language Generation Sydney, Australia.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284. http://dx.doi.org/10.1037/0033-2909.119.2.254

Linnebank, F., Liem, J., & Bredeweg, B. (2010). Question generation and answering. DynaLearn, Deliverable D3.3, EC FP7 STREP project 231526.

Liu, B. (2009). *SARAC: A Framework for Automatic Item Generation*. Paper presented at the Ninth IEEE International Conference on Advanced Learning Technologies (ICALT2009), Riga, Latvia.

Looney, J. (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: Organisation for Economic Cooperation and Development (OECD).

Luecht, R. M. (2013). An introduction to assessment engineering for automatic item generation. In M. J. Gierl & T. M. Haladyna (Eds.), *Automatic Item Generation*. New York, NY: Routledge.

McArthur, J., & Huxham, M. (2013). Feedback unbound - From measure to usher. In S. Merry, M. Price, D. Carless & M. Taras (Eds.), *Reconceptualising feedback in higher education*. New York, NY: Routledge.

McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment Research & Evaluation, 18*(2).

Merry, S., Price, M., Carless, D., & Taras, M. (Eds.). (2013). *Reconceptualising feedback in higher education*. New York, NY: Routledge.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review, 23*, 7-26.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner & R. Brunken (Eds.), *Instructional design for multimedia learning,* ,181-195). Munster, NY: Waxmann.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56-76. http://dx.doi.org/10.1016/j.compedu.2013.09.011

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*(2), 155-175. http://dx.doi.org/10.1207/s15326985ep2202_4

Nicol, D. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education, 34*(3), 335–352. http://dx.doi.org/10.1080/02602930802255139

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. http://dx.doi.org/10.1080/03075070600572090

Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education, 54*(3), 715-721. http://dx.doi.org/10.1016/j.compedu.2009.09.032

Papadopoulos, P. M., Demetriadis, S. N., & Stamelos, I. G. (2009). *The Impact of Prompting in Technology-Enhanced Learning as Moderated by Students' Motivation and Metacognitive Skills*. Paper presented at the 4th European Conference on Technology Enhanced Learning (EC-TEL 2009), Nice, France.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist, 40*(4), 235-244. http://dx.doi.org/10.1207/s15326985ep4004_5

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13. http://dx.doi.org/10.1002/bs.3830280103

Ras, E., Krkovic, K., Greiff, S., Tobias, E., & Maquil, V. (2014). Moving towards the assessment of collaborative problem solving skills with a tangible user interface. *The Turkish Online Journal of Educational Technology (TOJET), 13*(4).

Redecker, C., & Johannessen, O. (2013). Changing Assessment Towards a New Assessment Paradigm Using ICT. *European Journal of Education, 48*(1), 79-96. http://dx.doi.org/10.1111/ejed.12018

Richardson, J. T. E., Alden Rivers, B., & Whitelock, D. (2014). The role of feedback in the under-attainment of ethnic minority students: Evidence from distance education. *Assessment & Evaluation in Higher Education, 40*(4), 557-573. http://dx.doi.org/10.1080/02602938.2014.938317

Scriven, M. (1967). The methodology of evaluation. In W. D. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation , 1*, 39-83. Chicago: Rand.

Shepard, L. A., Hammerness, K., Darling-Hammond, L., Rust, F., Snowden, J. B., Gordon, E., Gutierrez, A.,Pacheco, A. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San-Francisco: Jossey-Bass.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189. http://dx.doi.org/10.3102/0034654307313795

Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning: The Journal of Open, Distance and e-Learning, 19*(1), 79-95. http://dx.doi.org/10.1080/0268051042000177863

Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324-328.

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2006). Classroom assessment for student learning: Doing it right-using it well. Portland, OR: Educational Testing Service.

Stödberg, U. (2011). A research review of e-assessment. *Assessment & Evaluation in Higher Education, 37*(5), 591-604. http://dx.doi.org/10.1080/02602938.2011.557496

Swan, M. (2012). Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Sensor and Actuator Networks, 1*(3), 217-253. http://dx.doi.org/10.3390/jsan1030217

Tabuenca, B., Kalz, M., & Specht, M. (2014a). "Tap it again, Sam": Harmonizing the frontiers between digital and real worlds in education. Paper presented at the IEEE Frontiers in Education Conference 2014: Opening Doors to Innovation and Internationalization in Engineering Education, Madrid, Spain.

Tabuenca, B., Kalz, M., & Specht, M. (2014b). Lifelong Learning Hub: A seamless tracking tool for mobile learning. Paper presented at the 9th European conference on Technology Enhanced Learning (EC-TEL 2014) - Open Learning and Teaching in Educational Communities, Graz, Austria. http://dx.doi.org/10.1007/978-3-319-11200-8_59

Tabuenca, B., Kalz, M., Ternier, S., & Specht, M. (2014). Stop and think: Exploring mobile notifications to foster reflective practice on meta-learning. *IEEE Transactions on Learning Technologies, 8*(1), 124-135. http://dx.doi.org/10.1109/TLT.2014.2383611

Tamin, R., Bernard, R., Borokhovski, E., Abrami, P., & Schmid, R. (2011). What forty years of research says about the impact of technology on learning: A second order meta-analysis and validation study. *Review of Educational Research, 81*(1), 4-28. http://dx.doi.org/10.3102/0034654310393361

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*(3), 323-342. http://dx.doi.org/10.1080/03054985.2012.689693

Whitelock, D. (2010). Activating Assessment for Learning: Are we on the way with Web 2.0? In M. J. W. Lee & C. McLoughlin (Eds.), *Web 2.0-Based-E-Learning: Applying Social Informatics for Tertiary Teaching* (pp. 319–342). IGI Global. http://dx.doi.org/10.4018/978-1-60566-294-7

Whitelock, D., Gilbert, L., & Gale, V. (2011). Technology-enhanced assessment and feedback: How is evidence-based literature informing practice? Paper presented at the International Computer Assisted Assessment Conference, Southampton, UK.

Whitelock, D., & Watt, S. (2008). Reframing e-assessment: adopting new media and adapting old frameworks. *Learning, Media and Technology, 33*(3), 153-156. http://dx.doi.org/10.1080/17439880802447391

Wiliam, D. (2008). Quality in assessment. In S. Swaffield (Ed.), *Unlocking Assessment: Understanding for reflection and application* (pp. 123-137). New York, NY: Routledge.

Wiliam, D. (2011a). *Embedded Formative Assessment*. Bloomington: Solution Tree Press.

Wiliam, D. (2011b). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14. http://dx.doi.org/10.1016/j.stueduc.2011.03.001

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy and Practice, 2*(5), 49-65. http://dx.doi.org/10.1080/0969594042000208994

---

[i] www.taotesting.com