British Journal of Educational Technology (2015)
doi:10.1111/bjet.12348

Assessment and learning technologies: An overview

Treasa Farrell and Nick Rushby

Treasa Farrell is education manager at the Council for the Curriculum, Examinations and Assessment in Northern Ireland. Nick Rushby has been working in the field of learning technology for over 44 years and is currently the editor of the British Journal of Educational Technology. Address for correspondence: Mr Nick Rushby, BJET Editor, Chantry Cottage, The Green, Otford, Kent TN14 5PD, UK. Email: nick.rushby@btinternet.com

Abstract

Assessment pervades the learning process. This paper provides an overview of the application of technology to support and enhance diagnostic, formative and summative assessment. The focus is on examining how it can replace what already exists, improve the functionality, catalyse a redesign of the process and in some circumstances, make possible what was previously inconceivable. The paper considers formal and informal individual learning environments, group settings and assessment at a national or international scale.

Practitioner Notes

What is already known about this topic

- A great deal has been carried on technology-enhanced assessment but the results are not widely known by learning technologists.
- Assessments based on multiple choice questions (MCQs) are difficult to devise but, especially when embedded in e-learning programmes, are assumed to be more reliable than they often are.

What this paper adds

- An overview of technology to support and enhance diagnostic, formative and summative assessment.
- A sense of different assessment types in the context of an increasing use of technology.
- A recognition that application of technology to assessment can be a major driver for change throughout teaching and learning.

Implications for practice and/or policy

- Think carefully about the purposes of their assessment and not just use technology-based assessment "because you can!"
- Think about how best to extract value from "big data" sets.
- Think about optimising opportunities for assessment within online learning.
- Think about how assessment can contribute to the optimisation of learning outcomes at national and/or international levels.

If a measurement matters at all, it is because it must have some conceivable effect on decisions and behaviour. If we can't identify a decision that could be affected by a proposed measurement and how it could change those decisions, then the measurement simply has no value (Douglas Hubbard, 2010).

Introduction

Assessment is the process of identifying, collecting and interpreting information about learning outcomes. It is an integral part of the teaching, training and learning cycle. The processes of assessment pervade the systematic application of learning, both in education and in training.

The classical systems approach, a general model of which is shown in Figure 1, starts with an assessment of what the learner (or learners) know or can do. After comparing this with the desired situation (in terms of knowledge, performance or behaviour), a learning intervention to close the gap between what is and what should be is developed. That intervention may take a variety of forms including, but not limited to, lectures, discussion, reading, project work—or e-learning. As the learner works through the intervention, it is likely that there will be some kind of formative assessment so the he or she can judge progress. That might take the form of responding to questions or reflecting on what has been learned. Then, particularly in formal learning environments, but less often in informal learning, there will be a summative assessment to judge what the learner now knows or can do. In the systems approach, that assessment then restarts the cycle, leading to another analysis of the gap between what is and what should be, and another intervention. And so the cycle is repeated.

Within that cycle there are other uses for assessment, feeding into the evaluation and (hopeful) improvement of the learning intervention. Some of the assessment information is used not for the direct benefit of the learner, but for the organisation so that it can determine how well it is doing—perhaps compared with similar organisations, or at a national or international level. So we have different types of assessment—summative (before and after the event) and formative (within the learning experience)—and a number of different stakeholder groups—the learners, the institution and the wider public.

The aim of this paper is to make sense of these different assessment types in the context of an increasing use of technology.

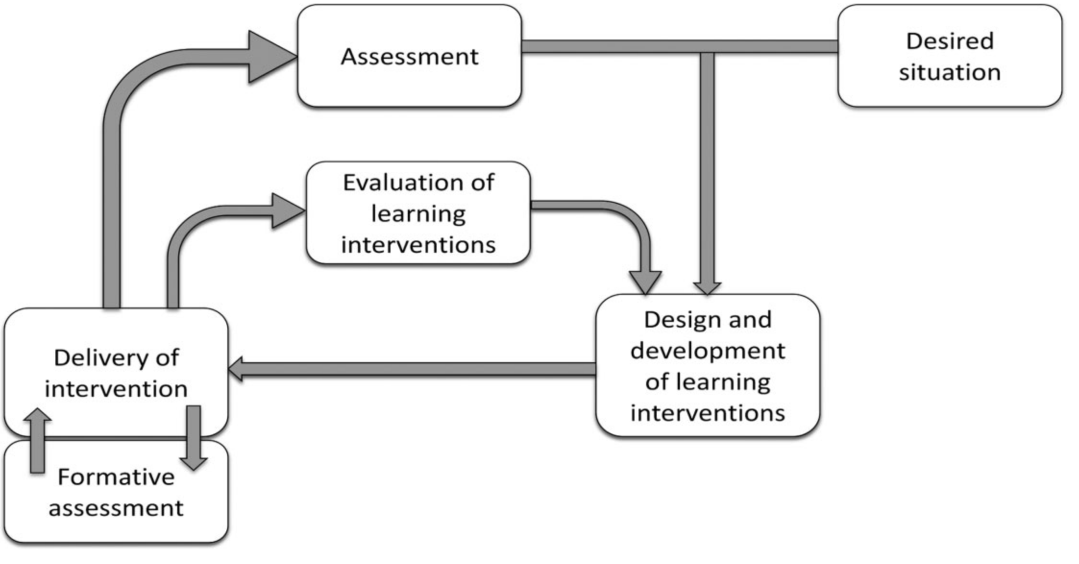

Figure 1: A systematic approach to learning

Some general thoughts about assessment

Assessment, of course, is nothing new. The Socratic Circle (Copeland, 2005) can be viewed as a means of teaching and learning through continual formative assessment. The Chinese Civil Service used written examinations for selection purposes dating back hundreds of years. Today, many nations assess young people to see whether they qualify for entrance to higher education and assess them at the end of their studies to see how well they have done. Children are assessed in schools (perhaps too frequently) to determine how "well" they are doing, to determine the "value" that the school is adding and, in the United Kingdom, to provide comparative data to help parents decide which school might be best for their own children. In business and industry, the emphasis is less on normative assessment (eg, determining who is in the top quartile or decile and arranging candidates in order of their performance) and more on criterion-referenced assessment (determining who has reached a given level of performance). Much of the assessment in business is for the purpose of demonstrating compliance. Technology has made it possible to collect and analyse much more assessment data than was possible using paper-based methods, and as technology is increasingly used to deliver learning, it seems natural to use that same technology to assess learning and to process the data for a variety of audiences.

This paper is based on the premise that assessment is an integral part of education and training and involves the measurement of what is important in the process. We must, however, be careful not to fall into the trap that what gets measured becomes important! As we shall see, technology-based assessment provides some important affordances. It can increase objectivity and reduces the resources needed to carry out assessment. In doing so, it increases the possibility that assessment will come to dominate the process. Abraham Kaplan said "Give a boy a hammer and everything he meets has to be pounded" (Horowitz, 1962). We should not assess just because technology makes it easier.

Whether or not technology is used to support assessment, we have to consider validity and reliability. This is not the place for a detailed discussion of these two aspects, but a basic introduction may be helpful because they impact on how technology is used (and perhaps, misused).

Validity concerns whether we are employing an appropriate measure. For example, if we want to assess an individual's understanding of geography, a series of questions about the life cycle of a butterfly is not likely to give us a valid measure! We have to ask whether what we are measuring is valid—or are we measuring it because we can, and then giving it undue importance. Reliability concerns whether our measurement gives repeatable results. For example, if we want to measure the length of a pencil, then an elastic band with centimetre marking is unlikely to give very repeatable, reliable results. Here, technology may give an advantage in delivering more repeatable results that human assessors (Tisi, Whitehouse, Maughan & Burdett, 2013).

Enhancement or transformation

Puentedura (2009) identified four levels through which we progress in our use of technology. The SAMR model (the acronym is taken from the initial letters of the four stages) takes us from:

1 Substitution, in which the technology is a direct substitute and there is no functional change, through
2 Augmentation, in which the technology is still a direct substitute but now with some functional improvement, to
3 Modification, in which the technology allows or even catalyses significant redesign of the tasks, and finally
4 Redefinition, where the technology enables us to create new tasks that were previously inconceivable.

As we introduce technology, we must continue to focus on what is it that we really want to assess. Where there is direct substitution (stage 1), there is no change in the assessment function; we are assessing the same things but more efficiently. The validity and reliability are unchanged. However, as we progress through the stages, there is the opportunity to redesign the assessment and perhaps get closer to measuring the right things—improving the validity as well as the reliability. Sometimes, because the technological affordances are not well aligned with our optimal assessment, there is a real risk of compromising the assessment. For example, multiple choice questions (MCQs) are easier to use with technology than short or essay questions, but they may not be ideal for our purpose. In contrast, it is arguable that assessing a learner's understanding of ecological concepts through the medium of a serious game could be a far more valid approach than traditional written examination questions.

MCQs

Within technology-based learning materials, there has been a move towards MCQs. These are more amenable to machine marking than short answer or essay-type questions and are often perceived to be more objective.

A typical MCQ consists of a stem followed by several alternative responses. Some of these are incorrect responses (distractors) while one or more are correct responses. The learner is required to select one or more of these alternatives depending on the form of the question. MCQs have a number of advantages:

1 They can assess different levels of learning of basic recall through to analysis and the evaluation of information.
2 They offer an objectivity that is difficult to achieve with the scoring of short answer or essay questions.
3 Because learners can answer the questions more quickly, the number of questions and the coverage of the curriculum can be increased. This has the effect of increasing the reliability and validity of the assessment.

There are several formats for MCQs, including:

- Choose the correct response from 4 to 5 alternatives.
- Chose from two responses (effectively a "yes/no" question).
- Choose several correct responses from a number of alternatives.
- Place a number of items into the correct order.
- Match the items in one list with those in a second list.

However, the formats place limits on what can be tested. It is difficult to devise MCQs that assess the learners' ability to organise their thoughts or to assess their creativity. It is also difficult to write good MCQs. An effective MCQ will always be answered correctly by those who know the subject material, and incorrectly by those who do not and will therefore discriminate between them effectively.

Common problems include stems that are ambiguous, and alternative responses where the correct choice is obvious. This leads to candidate strategies for answering MCQs that include eliminating alternatives that are clearly incorrect or that are overlong, and then guessing the correct answer from the remainder. In extreme cases (which are not uncommon), MCQs can be answered correctly by those who should fail the test, and incorrectly by those who should have passed.

Fortunately, the technology that is used to mark the questions can facilitate the analysis of the performance of the overall assessment as well as producing individual scores and rankings. By looking at how individuals answered each question, it is possible to determine whether some of the distractors (incorrect responses) are so unattractive that no one chose them. A disproportionate number of individuals choosing one specific distractor may indicate a systematic misunderstanding that could be attributable to a problem with the learning materials. If a

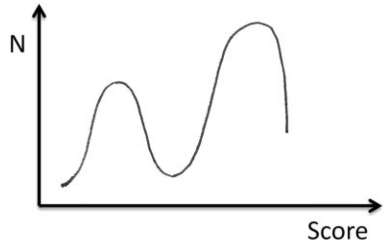

Figure 2: Candidates' scores for an "ideal" MCQ test in a training context

question is repeatedly answered incorrectly by individuals who otherwise have high scores in the whole test, then we can infer that it is ambiguous.

An ideal MSQ test in a training context might have a bipolar distribution of scores as shown in Figure 2. The high-scoring learners can be distinguished clearly from those with low scores.

In an educational context where the aim is set the learners in order of their performance, the test would include MCQs that ranged in difficulty to make the distribution more linear.

The effectiveness of typical MSQ assessments can be illustrated by use of the Skurnik–Nuttall Measuremeter (Skurnik & Nuttall, 1987). In summary, this analyses a test and gives an indication of how many distinct grades the test can allocate so that the majority of those taking the test will be in the correct grade, plus or minus one. So, for example, a test used to put individuals into one of seven grades "A" to "G," would require a Measuremeter score of 7. Even with this score, a significant number of individuals would be in the wrong grade—either one grade higher or one grade lower. For many MCQ tests, however, the score is often far smaller, indicating that they are ineffective at discriminating between more than two or three separate grades.

As the number of individuals taking the test increases, so the analysis of individual questions (the item analysis) becomes more precise. This enables ineffective and misleading items to be removed, and the reliability of the remainder to be improved by rewriting distractors. This can then result in greatly improved measurement scores.

Optical scanners were developed to read and process MCQ tests. The candidate marks the chosen response(s) with a black pen or pencil and, as the answer pages are scanned, the pattern of marks is detected. The pattern is compared with the rubric and a score is calculated. Although the scanners do not offer 100% accuracy, they are significantly better than can be achieved by human markers. They are also much faster.

With the growth of e-learning (computer-based learning), it became easier to present the assessment on the computer screen and ask the learner to select the correct response(s) online. Tests could be embedded in learning materials and used for formative assessment as well as for a final summative assessment. It was also easier to use other forms of MCQ such as ordering and matching as described above. However, the skills required to design good MCQs were not made redundant by the technology. A poor MCQ remains a poor MCQ whether it is administered on paper and marked by hand, or delivered and marked on a computer! Some attempts were made to use adaptive computing techniques to construct or select MCQ items in real time so that the difficulty of the assessment could be varied according to the learners' progress. With a few exceptions, these were not very successful and these remain in the research environment.

The problem of invigilation

One of the primary advantages of e-learning is that it can be available in any place at any time. This allows the learner to be freed from learning as part of a cohort in a lock-step environment.

Ideally, the summative course assessment should also be available to the learner at a time and place of his/her choosing. However, this freedom poses a major problem for ensuring that there is no cheating. How can you ensure that an individual actually taking an assessment is the person who is registered, and that he/she is not receiving any outside help?

A biometric approach for authentication was developed by Clarke, Dowland and Furnell (2013) while Ahlawat, Pareek and Singh (2014) describe a system for online invigilation that combines face recognition (with a webcam monitoring the candidate) and audio speaker recognition. Although they do not give statistics for the probabilities of failing to discover malpractices, it is quite possible that the presence of such a system will act as an effective deterrent to potential cheats.

Assessment for policymaking

A key purpose of assessment is to raise standards for the individual, for the entire cohort of an educational establishment, for an industry, for a country and internationally. Many parties have a vested interest in improving standards of education for a particular group and also across the world. Among these parties are the learners themselves; their parents or carers; the educational establishment which is providing for them; businesses who will be employing these learners in the future; and governments who are responsible for the overall economic and social well-being of a country or region.

At the strategic level, governments need to take account of the key economic and social drivers for their country. They need to consider the governmental aims for the development of the country; employment opportunities currently available and as they are predicted to be at various key times in the future; the potential supply of people with the right qualifications, skills and experience compared with the employment currently available and planned in the future; the current educational resources in the form of staff, buildings and equipment; the finance available; current curricula and assessments.

In order to promote economic success in a country, the government has to ensure that all its people are given opportunities to acquire skills and qualifications that will match with the employment prospects available. This also benefits employers as, “the ability of a company to exploit an idea is critically dependent on the availability and quality of human resources (numbers of people, skills, etc.) and capital resources (including finance and infrastructure)” (National Endowment for Science Technology and the Arts (NESTA), 2009).

Investment in technology is critical in strategic planning. It is vital that decisions are made based on educational principles rather than simply a desire to increase the use of technology in assessment.

> Education decision makers need to better align their investments to 21st century requirements. A significant increase in learning outcomes and an increased return on investment is possible but requires cost-benefit analysis as part of the transformation process. More collaboration is required to build on the most promising innovations in education, ensuring they can be scaled for the benefit of many more learners. Technology is a critical enabler to transform education aligned with new pedagogical frameworks and result-oriented investments (Weisschuh, 2012).

Assessment provides valuable information to governmental agencies. Governments worldwide are committed to a variety of assessments that are used for comparison with performances in other countries. Through the Organisation for Economic Co-operation and Development (OECD), governments work together for a range of aims, including providing a setting where they can compare policy experiences, seek answers to common problems, identify good practice, and work to co-ordinate domestic and international policies. Part of that activity involves assessments to enable comparisons of educational attainment (Exhibit 1).

The Programme for International Student Assessment (PISA) was set up by the OECD in 2000. This international tricentennial survey aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. To date, students representing more than 70 economies have participated in the assessment.

'PISA represents a commitment by governments to monitor the outcomes of education systems through measuring student achievement on a regular basis and within an internationally agreed common framework. It aims to provide a new basis for policy dialogue and for collaboration in defining and implementing educational goals, in innovative ways that reflect judgments about the skills that are relevant to adult life' (OECD, PISA, 2009, p. 9).

For the first 12 years, PISA used pen and paper tests because this allowed access to the widest range of learners across the world. However, from the 2015 event, computer-based assessment will be the default mode. The tests are marked in each country following the protocols set out and agreed by all participating countries and the results forwarded to PISA. At this stage, technology plays its part in analysing the data and providing valuable information to each participant country and to the worldwide educational community. This information plays a major part in the development of educational policies across the world. It allows educationalists to see where the strengths and weaknesses are worldwide and gives them the opportunity to amend education systems accordingly.

Exhibit 1: The Programme for International Student Assessment (PISA)

Class response systems

Obtaining timely feedback in a classroom environment presents problems for the teacher or instructor. Where the numbers are relatively small, it is possible to watch for the non-verbal signs that indicate a learner is struggling, bored or off the pace of the session. As the class numbers rise, this becomes more difficult. Group response systems, first developed in the 1980s, provide a means of closing the distance between the teacher and the learners.

The basic idea is straightforward: each learner has a small device, about the size of a small mobile phone, with a numeric keypad (sometimes with some additional yes/no keys). When asked by the teacher, the learner presses one of the keys to indicate their response to a particular question. The individual responses are transmitted to a computer and analysed to provide a display for the teacher. While early systems used physical cables to connect the response units to the computer, the wires have now been replaced by infrared or wireless communications.

The group response systems can be used to check that all of the learners have understood a particular point by answering an MCQ, or to take instant polls of their views on certain points. The software running on the PC often enables immediate displays of the results visible to the whole class, or can highlight non-responders on a screen visible only to the teachers.

There are a number of obvious logistical problems, not least the propensity for the number of response units to decrease with each session, as they get lost or mislaid. There is also the necessity of registering units to specific learners.

The assessment of competence

The growth of National Occupational Standards in the 1990s heralded a move towards more objective assessment in vocational training within the United Kingdom (UK Commission for Employment and Skills, 2014). This approach has now been adopted more widely in Europe and elsewhere. The standards are based on the concept of competence, which involves the assessment of whether an individual can perform a specified task under specific conditions, to a specified standard. This is an objective test of performance with candidates being required to demonstrate that they can perform a certain task.

The assessment of competence in the workplace has moved to the heart of many effective training systems. Companies in the rail, petrochemical, offshore, nuclear and aerospace industries need the means to assure themselves that their staff are competent to carry out safety critical work. This often involves complex equipment, a highly dispersed workforce, unpredictability of work patterns and the need for spontaneous assessment, often within a hostile environment. Companies in banking and finance need assurance that their staff comply with statutory requirements. The work environment may be more benign, but the penalties for failure are high. An accurate understanding of what staff know and can do now, and of what they need to know and be able to do, enables training to be focused more accurately and at the appropriate time, resulting in substantial cost savings through reduced wastage. However, unless that assessment is managed carefully, it becomes an impossible burden. A significant improvement in productivity is required if these goals are to be achieved within reasonable costs and technology holds a key to that increased productivity.

The move to competence-based assessment can result in a large amount of data which needs to be effectively collected and managed. The assessor can be confused as to what has already been assessed and what is still to be completed. This opens up possibilities for technology support. One approach is described by Rushby (1996) (Exhibit 2).

British Rail (the UK infrastructure owner, now Network Rail) was chosen as a specific example of an organisation needing to carry out workplace assessment in a hostile environment. At a detailed level, there were some 4000 elements of competence for the technicians carrying out first-line maintenance and faulting of the signalling system. Each competence would typically consist of 6–10 performance criteria which needed to be assessed against defined criteria under a range of circumstances. Each technician would require a subset of these competences.

The project was able to estimate the costs that might be expected for a Maintenance Unit with 1000 technicians in signal and telecommunications, civil, mechanical and electrical engineering, and with 100 supervisors, each carrying a Newton MessagePad. This indicated that there would be 30% financial savings in the assessment process. However, the major cost saving for the railway and for other organisations where competence for safety critical work is an issue, is in reducing the risk of losses through *lack* of competence. An independent risk analysis, carried out in 1992 by the British Rail's own risk assessors, estimated that if "competence" in signal and telecommunications could be increased by 10%, this would reduce the losses by up to £65 million (about US$104M) per year (at today's prices). This potential saving would dwarf the costs of introducing a system of workplace assessment supported by PDAs.

The project used the Apple Newton MessagePad (PDA) to guide the assessor through the maze of assessments so that they could take advantage of opportunities in the workplace, and to record the results. These would later be uploaded to the training database. Now, 20 years later, the task could be handled by a smartphone, but the underlying concepts remain the same.

Exhibit 2: Workplace assessment in hostile environments

Other systems (eg, PaperFree) are described by Whitelock, Mackenzie, Whitehouse, Ruedel and Rae (2006).

Assessing essays and free-form responses

The analysis of natural language text presents problems of a much higher level of difficulty than MCQs. Let us consider first questions where the learner is asked to type a few words in response. It is relatively simple to recognise key words and phrases and determine whether they are present. If the order is significant, then this, too, can be determined. There needs to be some check on whether a negative is present. The response "This is not a parrot" is clearly different from "This is a parrot." Beyond this, it gets much harder but is still achievable! More information on this topic can be found in Clark, Fox and Lappin (2012), Jackson and Moulinier (2007), and Collobert *et al* (2011).

The difficulty of analyzing natural language responses poses three questions:

1 How reliable is automated assessment of free-form responses?
2 How is it perceived by learners?
3 How is it perceived by teachers and instructors?

In looking at the reliability, we have to accept that the bar is not set very high. Research into the reliability of human assessors (reviewed in Tisi *et al*, 2013) considers a number of factors which impact on reliability including:

- Professional expertise (so that expert assessors are better at interpreting responses than those who lack subject expertise);
- Systematic leniency or severity (which can be dealt with by moderating the marks);
- Inconsistent marking (which cannot be adjusted for and, if detected, requires re-marking).

In practice, their review indicates that human assessors are not very reliable and for high-stakes assessment (eg, summative assessment at important points in the course), procedures have to be put into place to moderate, check and trigger remarking, so that reliability is improved.

In contrast, assessment by computer is at least consistent and, because it can collect large amounts of data on its performance, it is much easier to improve its reliability. Jordan and Mitchell (2009) found that "A computerised system has been shown to accurately mark short-answer free-text questions and to deliver tailored feedback on incorrect and incomplete responses." Tisi *et al* (2013) note that

> There have been some advancements in the field of computer-based marking. However, with the exception of some objectively marked item types, much work is still needed before computer marking becomes a viable alternative to human marking. . . . In addition, technological advances have facilitated the measurement of marking reliability and led to an increase in our understanding of the influencing factors. This, in turn, has produced tangible methods for improving marking reliability that can be implemented in high stakes examinations in England.

Despite the evidence concerning objectivity and increased reliability, it seems that both the teachers and the learners do not entirely trust automated assessment of essay and free-form responses (Lai, 2010). In consequence, their use has been mainly in low-stakes assessment (formative assessment) rather than in high stakes, summative assessment. However, learners *do* appreciate the improved feedback that can be provided by the same system. Although the use of technology is a substitute for the human assessor and the function of the assessment is not really changed, it improves the efficiency of the process. In many circumstances, the human tutor would simply have insufficient time to provide the level of detailed feedback that is valued by the learners.

Those working in the field of artificial intelligence in the late 1960s and 1970s took an interest in techniques that could be used to assess a learner's performance in more complex tasks and where the answers to questions might be given in a variety of forms. Human marking had little difficulty in judging whether these responses indicated a correct (but possibly misspelled) answer, or whether they were just wrong. In the middle of the 1970s, attention turned from trying to establish what was "right" to indentifying what might be "wrong" in incorrect answers. This required the design and development of sophisticated models of the learner in which procedural errors could be diagnosed as the correct execution of an incorrect procedure, rather than fundamental misconceptions. This line of research was exemplified by the BUGGY research described in Exhibit 3.

> BUGGY was an example of an artificial intelligence approach to teaching algebra (eg, the subtraction of multi-digit numbers) by focusing on the *errors* that the learners made (that is, the "bugs" in their mental procedures). It was developed in 1975 by Brown and Burton (Brown & Burton, 1978; Burton, 1982). They designed a diagnostic model which reflected the learners' understanding of the skills and sub-skills involved in a task. BUGGY then used its student model to simulate a student with "buggy" thinking and generated diagnostic assessments to identify students' mistakes (eg, the student subtracts the smaller digit in each column from the larger digit regardless of which is on top).
>
> This diagnostic approach was a milestone in the research on student modelling and technology-based learning. It provided researchers and teachers with a specific diagnosis of a student's knowledge and capabilities on specific sub-skills. Although it required considerable resources to build the initial repertoire of incorrect procedures, its diagnoses appeared to be extremely helpful.

Exhibit 3: Buggy

Assessment and reflection

It is a general tenet of learning that effective learning programmes provide continual opportunities for the learners to check their understanding and monitor their progress. Within e-learning, we find frequent use of MCQs for formative assessment to provide feedback and for diagnostic feedback to the programme itself so that it can adapt to the individual's successes and problems. At a higher level in the overall programme, learners are encouraged to reflect on what they have learned and how they have learned it. This exercise might involve a learning journal; the learner keeps a diary with notes, thoughts and observations with the aim of enhancing his/her learning through the process of writing and thinking about the experience. Increasingly, that diary is likely to be maintained as a digital document.

Taking this a step further and adding structure leads us to the concept of the e-portfolio. Unlike the learning journal, this document (or collection of documents) is intended for wider publication, for example to support an application to a higher education institution or for employment (see Stefani, Mason & Pegler, 2007). Tzeng and Chen (2012) found that students were more inclined to create and use e-portfolios when they perceived their use for potential employers. However, this dual use—for formative assessment and job seeking—can create tensions. For example, Barrett and Carney (Barrett & Carney, 2005; Barrett, 2007) found that using e-portfolios for formal assessments can be an impediment to constructivist learning.

A reflective blog can be used to share the learning journal with other course participants. At this point, the boundaries between learning and teaching become increasingly blurred. Garcia, Elbeltagi, Brown and Dungay (2014) found evidence to suggest that:

. . . the use of blogs results in a learner role which is more focused toward seeking and providing peer critique, support and guidance, and as a result, there is a need for learners to be fully engaged and be willing participants within group learning. The learner role when using collective blogs therefore appears to meet the requirements of a connectivist learning environment as the learners' role becomes increasingly concerned with self-management, knowledge management and network building within the context of the blogging community, and the need to take greater responsibility for individual learning, which does not necessarily occur naturally.

Assessment and the disadvantaged learner

An essential part of planning for assessment is to allow for fair and equitable assessments for all learners. Most countries have legislation which requires that reasonable accommodations be made to the way that learners' education is provided to cover any disabilities. The legislation usually requires learners' needs to be anticipated—so any necessary adjustments should be in place before they are actually required. There are implications for assessment, but the actions you might take to anticipate the learner's assessment needs will often benefit all learners. However, it may be that if all learners are not considered fully when assessments are created, then the learning technology may become a barrier to equal opportunity.

The use of technology in assessment can allow those with physical difficulties to access assessments in multiple formats and to complete them in a similar time frame to their peers or to allow for more time when needed. Technology can also allow greater independence and so improves self-esteem. For example, consider an interactive computer-based test that relies on spoken instructions delivered through a headset. A learner with a hearing impediment would be severely disadvantaged. To counteract this disadvantage, the instructions could also be available through a video where the instructions can be lip read and also signed. Care is needed with written instructions because, in some circumstances, these might require a higher level of reading ability than the rest of the test. Questions where time is a factor might also need to be adjusted to allow for the differing time that it may take to receive and understand instructions delivered in this alternative form.

However, the quest for a level playing field must also take into account those learners who have difficulty in using the technology itself. And where assessments are used across countries, there must be consideration of whether all of the candidates have access to the technology. Assumptions that everyone has access to the Internet and a computer are not valid for many developed countries and certainly questionable in developing countries.

New assessment models must not erode efforts to promote high expectations for all students; nor should they disadvantage low-income schools and students with currently limited access to technology (Tucker, 2009).

Technology-based assessment challenges our assumptions as to who is disadvantaged—and how.

Technology and the management of assessment

Formative assessment of learning is a labour-intensive process. Teachers and instructors have traditionally, on a regular basis, given written or verbal feedback that is constructive and allows the learner to not only see their mistakes but to be given strategies to improve. Lefevre and Cox (2014) note that "feedback has a powerful effect on learning" and have found that sophisticated and differentiated types of feedback improves that effectiveness. That takes time and effort which is not easily found when the emphasis is on improving quality while reducing costs.

Reliability with human markers is problematic unless there are layers of checking and moderation. These add to the assessment task. And the task of marking summative assessments to tight deadlines can be daunting where large number of candidates are involved.

As we have seen throughout this paper, assessment is pervasive. It is encountered throughout the learning cycle for the purposes of diagnosis, giving formative feedback and in determining the success of the learners—and the learning process. In some contexts, it is embedded so deeply into the learning that it is difficult to distinguish assessment activities. It is therefore an obvious step to use technology to help manage the various assessment activities and to collate the plethora of data that are available.

The UK Joint Information Systems Committee (now known as JISC) funds an assessment and feedback programme which is focused on supporting large-scale changes in assessment and feedback practice, supported by technology, to efficiencies and quality improvements. One of the outputs from that programme is a report on the Electronic Management of Assessment. That report (which is written primarily from mainly a higher education context) identified a number of benefits for students where there is an integrated process including:

- Convenience of not having to travel to hand in assignments;
- Avoidance of printing costs;
- Time savings and avoidance of anxiety about assignments going missing in the postal system (in distance learning environments);
- Automatic proof of receipt with time-stamping;
- Improved confidence provided by the privacy, safety and security of e-submission;
- Confidence of knowing work is backed up;
- Electronic reminders about deadlines and improved clarity about turnaround times for marking;
- Improved clarity and understanding of feedback (not least as a result of not having to decipher handwriting);
- Improved timeliness of feedback (especially when some aspects are automated) enabling advice given on a previous assignment to be assimilated and applied in the next;
- Realistic timing of submission deadlines;
- Meeting expectations—normal practice in a digital age;
- Increased privacy when marked work is returned electronically.

For staff, the reported benefits included:

- Greater transparency which has been shown to improve the standard and consistency of marking and feedback comments
- Improved clarity of marking and feedback (especially the ability to include lengthy comments at the appropriate point in the text). This also increased morale through not having to write out repeated comments
- Reduced workload making it feasible to assess learners' understanding more frequently and to focus on those individuals with difficulties
- New opportunities to improve student understanding, for example, by extracting and analysing data held in an online marking system to achieve a more timely response to errors and weaknesses.
- Increased satisfaction when improved feedback has a positive impact on student attainment
- The convenience and security of having assessment information stored and backed up electronically
- The ability to moderate marks without having to physically exchange paper
- The increased speed and efficiency of being able to reuse common comments
- The convenience of being able to undertake originality and plagiarism checking in the same environment as marking
- Reduced data input and batch upload of marks (Ferrell & Gray, 2014).

Learning analytics

The widespread use of technology in all aspects of the assessment process for diagnostic assessment, monitoring the learners' through formative assessment and in high-stakes summative assessment can provide an immense amount of data on the learners, their teachers or instructors, and the learning interventions with which they engage. This has led to a growing interest in learning analytics as a means of understanding and improving the learning process and the environments in which learning occurs. Ferguson (2012) identifies three key drivers and three associated questions for learning analytics:

- The growth in big data: How can we extract value from these big sets of learning-related data?
- The rise of online learning: How can we optimise opportunities for online learning?
- Political concerns for improvements in education: How can we optimise learning and educational results at national or international levels? This resonates with the earlier discussion about international comparisons.

Research in learning technology has too often been characterised by small-scale studies involving a relatively few subjects and focusing on a limited number of parameters. The tools that are being developed for big data enable us to combine and extend these datasets while the new analytical tools give us ways to make sense of the data—to extract useful information that will enable us to understand what is happening and how to improve it.

Conclusion

At the start of this paper, we used Puentedura's (2009) model to identify four levels through which we progress in our use of technology. We conclude with a some thoughts on how far different applications of technology to the assessment process have progressed.

Table 1 gives our personal view of the impact of technology on assessment. Other commentators might disagree as to how far we have progressed!

The view that electronic management of assessment is essentially using the technology to automate the processes involved should not be taken to minimise the impact that this can have on the resources (and thus the costs) required. It could be argued that if those saved resources can be directed towards improving feedback and helping learners with difficulties, then it should be considered to be augmenting the process.

Certainly, the use of technology to assess free responses, and thus increase transparency and objectivity, is providing a functional improvement.

We suggest that the use of MCQs (at least, the more sophisticated forms of MCQ) demonstrates a significant redesign of the task and lifts this application to the level of modification. The same can be said of the affordances of reflective blogs and e-portfolios.

And finally, we have three applications of technology that have enabled us to create new tasks that were previously inconceivable. The group response system is a simple idea which enables a redesign of large group teaching methods, while learner modelling and learning analytics would be impossible without technology.

Table 1: The impact of technology-based assessment

	Substitution	*Augmentation*	*Modification*	*Redefinition*
Electronic management of assessment	✓			
Computer essay marking		✓		
Multiple choice questions			✓	
Reflective blogs			✓	
E-portfolios			✓	
Group response systems				✓
Modelling learners' problems				✓
Learning analytics				✓

Because assessment pervades the learning process, the application of technology to assessment can be a major driver for change throughout that process.

References

Ahlawat, V., Pareek, A. & Singh, S. K. (2014). Online invigilation: a holistic approach: process for automated online invigilation. *International Journal of Computer Applications*, *90*, 17, 31–35.

Barrett, H. & Carney, J. (2005). *Conflicting paradigms and competing purposes in electronic portfolio development*. Retrieved December 18, 2009, from http://electronicportfolios.org/portfolios/LEAJournal-BarrettCarney.pdf.

Barrett, H. C. (2007). Researching electronic portfolios and learner engagement: the REFLECT initiative. *Journal of Adolescent & Adult Literacy*, *50*, 436–449. doi: 10.1598/JAAL.50.6.2.

Brown, J. S. & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, *2*, 155–191.

Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. H. Sleeman & J. S. Brown (Eds), *Intelligent tutoring systems* (pp. 157–182). London: Academic Press.

Clark, A., Fox, C. & Lappin, S. (Eds) (2012). *The handbook of computational linguistics and natural language processing*. Malden, MA: Wiley-Blackwell. ISBN-10: 1118347188.

Clarke, N. L., Dowland, P. & Furnell, S. M. (2013). e-invigilator: a biometric-based supervision system for e-assessments. International Conference on Information Society (i-Society). IEEE. pp. 238-242.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*, 2493–2537.

Copeland, M. (2005). *Socratic circles: fostering critical and creative thinking in middle and high school*. Portland, Maine: Stenhouse Publishers. ISBN-10: 1571103945.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*, 5/6, 304–317.

Ferrell, G. & Gray, L. (2014). *Guide to the Electronic Management of Assessment*. Bristol, UK: JISC. Retrieved November 3, 2014, from http://www.jisc.ac.uk/guides/electronic-assessment-management

Garcia, E., Elbeltagi, I., Brown, M. & Dungay, K. (2014). The implications of a connectivist learning blog model and the changing role of teaching and learning. *British Journal of Educational Technology*, *46*, 4, 877–894.

Horowitz, M. J. (1962). Trends in education (report on the annual meeting of the American Educational Research Association held on February 19–21, 1962). *Journal of Medical Education*, *37*, 790–793.

Hubbard, D. W. (2010). *How to measure anything: finding the value of "intangibles" in business*. Malden, MA: Wiley.

Jackson, P. & Moulinier, I. (2007). *Natural language processing for online applications: Text retrieval, extraction and categorization* Vol. 5. Philadelphia, PA: John Benjamins Publishing.

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*, 371–385. doi: 10.1111/j.1467-8535.2008.00928.x.

Lai, Y.-H. (2010). Which do students prefer to evaluate their essays: peers or computer program. *British Journal of Educational Technology*, *41*, 432–454. doi: 10.1111/j.1467-8535.2009.00959.x.

Lefevre, D. & Cox, B. (2014). Feedback in technology based instruction: learner preferences. *British Journal of Educational Technology*. doi: 10.1111/bjet.12227.

National Endowment for Science Technology and the Arts (NESTA) (2009). *The wider conditions for innovation in the UK: How the UK compares to leading innovation nations*. Retrieved October 27, 2014, from http://www.nesta.org.uk/publications/wider-conditions-innovation-uk

OECD, PISA (2009). *Assessment framework: key competencies in reading, mathematics and science*. Paris: OECD Publishing. Retrieved July 29, 2015, from www.oecd.org/pisa/pisaproducts/44455820.pdf.

Puentedura, R. (2009). *SAMR: a contextualised introduction*. Retrieved October 5, 2014, from http://www.hippasus.com/rrpweblog/archives/2014/01/15/SAMRABriefContextualizedIntroduction.pdf

Rushby, N. J. (1996). Assistants for assessment. *Innovations in Education and Training International*, *33*, 3, 154–161.

Skurnik, L. S. & Nuttall, D. L. (1987). *Examination and item analysis manual*. Nottingham, UK: University of Nottingham. ISBN-10: 0853591334.

Stefani, L., Mason, R. & Pegler, C. (2007). *The educational potential of e-portfolios*. Abingdon, UK: Routledge. ISBN 978-0-415-41214-8.

Tisi, J., Whitehouse, G., Maughan, S. & Burdett, N. (2013). *A review of literature on marking reliability research (Report for Ofqual)*. Slough: NFER.

Tucker, B. (2009). *Beyond the bubble: Technology and the Future of Student Assessment*, Education Sector at American Institutes for Research. Retrieved October 31, 2014, from http://www.educationsector.org/publications/beyond-bubble-technology-and-future-student-assessment

Tzeng, J.-Y. & Chen, S.-H. (2012). College students' intentions to use e-portfolios: From the perspectives of career-commitment status and weblog-publication behaviours. *British Journal of Educational Technology*, *43*, 163–176. doi: 10.1111/j.1467-8535.2010.01165.x.

UK Commission for Employment and Skills (2014). *National Occupational Standards*. Retrieved October 24, 2014, from http://nos.ukces.org.uk/Pages/index.aspx

Weisschuh, J. (2012). Proposal—Guiding principles for investors to transform education for the 21st century, Global Economic Symposium. Retrieved July 15, 2015, from http://www.global-economic-symposium.org/knowledgebase/the-global-economy/effective-investments-in-education/proposals/guiding-principles-for-investors-to-transform-education-for-the-21st-century

Whitelock, D. M., Mackenzie, D., Whitehouse, C., Ruedel, C. & Rae, S. (2006). Identifying innovative and effective practice in e-assessment: findings from seventeen UK case studies. In M. Danson (Ed.), *Proceedings of the 10th CAA international computer assisted assessment conference* (pp. 505–511). Loughborough, UK: Professional Development, Loughborough University.