REVIEW PAPER 12: The Use of Technology for Assessment

Heidi Leeson & John Hattie

Visible Learning Labs

Faculty of Education

University of Auckland

Introduction

In most areas of society, technology is playing a central role on changing existing behaviours, functions and practices. The education space is no different. Not only is an understanding of technology becoming a substantive requirement for successful functioning in the workplace, technology is becoming a substantial medium for supporting and enhancing the learning process and teaching environment (Bennett, 2002). Classrooms are now utilising technologies ranging from simple tool-based applications (e.g., word processing, spreadsheets) to online resources (wikipedias, blogs, learning media, Moodle) that are accessed via various hardware components such as computers, laptops, and mobile handheld devices (e.g., PDA's, IPods).

Similarly, the effect of technology relating to what is taught in the classroom has extended to how this learned information is assessed. The impact of technology on assessment has seen such innovative practices as paper-and-pencil tests transferred to screen, automated scoring (e.g., of essays), item presentation adaptive to ability levels, items design that can be dynamic, adaptive, and interactive, automatic item generation, and boundary-less beyond classroom testing. In addition to current uses, future developments may see assessments conducted in a virtual reality where the student is experiencing or participating in a simulation of an event or problem, from within which they give a 'virtual' response. Real-time manipulation of screen entities and objects (e.g., within an interactive Science lab), would permit tracking of the steps taken to answer a problem or produce a result.

Thus, it is within both the current and future applications of technology to assessment that this paper will be focused. Technologies will be presented in two categories, those relating to technological in software applications, and those specific to hardware devices and componentry. Where it exists, research will be presented outlining the educational outcomes that are derived from the synthesis of these technologies. Issues and implications relating to the application of technological approaches to these areas of assessment will be explored, drawing particular reference to international research and evaluation reports.

Where possible, the following sections are presented in relation to both learning and assessment issues. However, there are some areas where technologies whilst fully applied to learning approaches, have not yet been extended to the assessment of that learning.

Hardware Devices and Componentry

Assessment Delivery – Hardware Devices

Technology is playing a vital role in enabling data-driven decision-making to occur. For example, web-based reporting systems can allow information relating to state and national curriculum standards to be aligned with the raw data collected from classroom assessments (common in the USA). In addition to these reporting systems, the use of mobile computing devices provide teachers with a platform to both administer and analyse data from classroom assessments against learning objectives. Schools have seen their technology infrastructure become more developed, particularly in relation to the student-to-computer ratio. Linked to the availability of wireless networks and funding, many schools are moving to mobile computing resources for teachers and students. Continued improvements in connectivity and access to the Internet will allow hardware teaching tools, like interactive whiteboards, to have more of a mainstream presence in classrooms. This section investigates the practices and research surrounding the move towards the mobilising of computer-based tools, and the effects that such devices appear to have on learning and assessment.

Given that hand-held computer devices now provide computing performance similar to previous generation laptops, hand-held devices are being seen as a viable alternative to delivering educational assessment (Trinder, Mahill, & Roy, 2005). Since its arrival onto the market 1994, Personal Digital Assistants (PDAs) technology has arguably made the most significant impact on the hand-held industry. PDAs (Palm and PocketPC) have other distinct advantages over PCs and laptops. There is no required boot up time (switch on and use) associated with PDAs, making them ideal to utilise at a moments notice. Further, unlike laptops, PDAs battery lifetimes are measured in weeks, rather than hours. However, in addition to the obvious advantages gained from portability, storage capacity, and usability, the price reduction of PDAs have propelled their use across schools, closing the 'digital divide' between well-funded and less well-funded schools. Their small size makes them a potable and non-intrusive piece of hardware to use for assessment (that is, where discrete observations and response recording is required).

Evidence from various educational-technology projects and initiatives (e.g., ICT Test Bed Project; The Learning Hubs Project) indicate that the implementation of electronic systems has resulted in substantial improvements in the detail and efficiency of both assessment and target setting practices (Somekh et al., 2007). For example, The Learning Hubs Project, involving 58,000 students and 5,000 teachers, sought to integrate technological advances of mobile computer devices, in order to adjust the teaching to reflect the use of 'real world' technology as found in the working environment. In addition, the aim of the project was to improve educational attainment, student engagement, and pedagogic performance of students by providing both teachers and students with a personal learning device, typically a PDA or Tablet. Results from this project showed that the one-to-one learning provided by the hand-held devices vastly improved student motivation, engagement and learning attitude. Teachers reported that the engagement and interest shown by the students resulted in a more enjoyable, creative, and productive classroom environment (Somekh et al., 2007). In addition, teachers preferred the mobility provided by the hand-helds over desk-bound PCs or power-bound laptops. To use Renaudon-Smith's (2007) analogy "who would build a room full of calculators for Maths".

In line with these findings research conducted by Becta (2007)[1] has found that learners show a deepening of engagement due to their increase involvement and interaction. Participant feedback has also indicated that students are typically more comfortable engaging in personalised or private subject areas using a hand-held mobile device, such as a PDA, than doing so using traditional methods.

An excellent example of the interaction of software and hardware technology is demonstrated by market-leader Wireless Generation, a United States for-profit company (http://www.wirelessgeneration.com/). Their hand-held computer (PDA) is loaded with the specific mCLASS software platforms (e.g., Dynamic Indicators of Basic Early Literacy Skills: DIBELS) depending on assessment requirements. This platform guides teachers through the administration of the assessment, and allows for assessment information to be recorded simultaneously by the teacher as the student completes each aspect of the assessment tasks. Performance information is then uploaded and synchronised with pre-existing information (e.g., previous performance benchmarks, state learning objectives/standards) thus, providing analysis (e.g., response pattern tracking), student- and item- level reporting, and individual and group level progress monitoring. Obviously, the development of this type of assessment technology has greatly enhanced the flexibility that teachers as test administrators have when collecting data. The Visible Learning Lab at the University of Auckland has already converted some of the asTTle functionality onto PDAs.

Software Applications

Assessment Design – Innovative Item and Test Design

Development of powerful computers and technological advances in communications is enabling the use of a new range of media, such as the World Wide Web and virtual realities. Such media is reshaping both face-to-face and computer-based learning and assessment with in education. The availability of multimedia and hypermedia technologies permit learner-controlled interactive solutions, where multiple media formats (text, video, audio, still images) can be displayed simultaneously. This section will review some of the technological advances that have been applied to innovative items and test design (e.g., multimedia,

[1] Becta is a government organization charged with developing ITC in schools across the United Kingdom

computer-based simulation), in addition, possible innovative approaches discussed (e.g., virtual reality).

To date, the term *innovative item* or *innovative test design* has encompassed a broad and diverse range of designs which often, only share as their commonality, the fact that they are computerized. As such, the term innovative item has been applied to such items ranging from essentially static passage/text editing (e.g., Breland, 1999; Parshall, Davey, & Pashley, 2000) to interactive video (IAV) computerized assessment (Dyer, Desmarais, Midkiff, Colihan, & Olson, 1992; Drasgow, Olson-Buchanan, & Moberg, 1999). This diversity indicates that the term is highly generic, encompassing designs that differ vastly in terms of presentation and test-taker interactivity.

As previously mentioned the capabilities of computerized testing present numerous opportunities for test developers to include various graphical aspects into items (Drasgow & Mattern, 2006). To date, however, only a few published studies present tests that have been created using a combination of multimedia features such as integrated audio, computer graphical technology and video. The most closely aligned innovative approach to the interactive scenario-based items in this study is the IAV computerized assessment (e.g., Desmarais, Dyer, Midkiff, Barbera, Curtis, Esrig, & Masi, 1992). IAV's items consist of video and sound, whereby the test-taker watches a video clip and at certain points during the clip is asked questions regarding the construct of interest. Certain items throughout the test are adaptive in that the test-taker's response dictates which video clip will be next presented (Drasgow et al., 1999).The most early example of an IAV assessment is the Workplace Solution test (Desmarais et al., 1992; Desmarais, Masi, Olson, Barbera, & Dyer, 1994; Dyer; Midkiff, Dyer, Desmarais, Rogg, & McCusker, 1992) where situation judgment skills within the workplace were assessed using a series of 30 scenes set in the context of a fictional organization.

Another organizational tool, Allstate Multimedia In-Basket (Ashworth & McHenry, 1992) used video clips representing the context of an airline customer service department to assess potential applicants' ability to conduct various tasks and display key knowledge and skills identified for the position. In addition to videoed instructions from the fictional "boss", test-takers interact with various graphics onscreen (e.g., filing cabinets), in order to complete the required tasks (McHenry & Schmitt, 1994). Unfortunately, no psychometric information is available for either the Workplace Solution test or the Allstate Multimedia In-Basket assessment applications.

One of the most recent and sophisticated uses of the IAV approach was in the development of the Conflict Resolution Skills Assessment (CRSA; Olson-Buchanan et al., 1998). Based on the KO (Keenan & Olson, 1991) model of conflict resolution, the stimulus for this measure consisted of the presentation of a scene where a conflict was taking place (typically 1-3 minutes in duration). At significant junctions during the scene, the test-taker was presented with several options for resolving the conflict, and based on the response given, the conflict scene continued to unfold. Thus, the IAV is adaptive in that certain scenes progress based on the test-taker's response to the previous video clip. Whilst the internal consistency of the test was high (r = .85), and no adverse impact was found amongst test-takers, only a modest correlation (r = .14) was found between the criterion (derived from a composite of on-the-job performance ratings from supervisors) and the CRSA (Olson-Buchanan et al., 1998). Thus, whilst Olson-Buchanan et al. acknowledged that further work was required to understand the results found in this study, especially in relation to validity, they did argue that the increased authenticity created through the use of the interactive video aspect was an important design concept to pursue and develop.

Leeson (2008) explored the use of interactive multimedia assessment, with particular reference to showing the additional measurement information than could be gained from this method compared to pencil-and-paper methods. The students engaged in a web-based multimedia (integration of both graphics and video) tool featuring a scenario whereby onscreen characters had to complete a challenging reading test. Throughout the scenario one of the onscreen students turned to the user and asked – "Is this how you feel?" or similar, in reaction to a vignette played out in the scene. There was higher levels of information (in the measurement sense) for the students with lower proficiency (in reading self-efficacy), showing how the interactive multimedia medium can lead to more dependable measurement with less negative social desirability influences compared to the pencil-and-paper method (Leeson, 2008).

An exciting direction for innovative assessment relates to the use of virtual environments, like those found on websites such as Second Life, IMVU, There, Active Worlds, and Kaneva. Much of the research and development that has been conducted at applying a virtual reality to the educational realm has been focused on the Second Life platform. Opening to the public in 2003, Second Life (Linden Research, Inc) is a 3-D virtual world, consisting of millions of square meters of virtual lands, and residents (avatars) which interact with each other (and objects) in a cyber-social network environment. Universities from around the world have utilised the virtual environment for hosting various initiatives, such as conducting discussion groups, delivering lectures, and hosting and managing projects and forums (Lagorio, 2007). The schome-NAGTY (The National Association of Gifted and Talented Youth) Teen Second Life Pilot explored the learning acquired and evidence of learning approaches used by adolescence in a Second Life island (Schome Park). Results indicated that there was strong evidence of high engagement levels amongst the students within the Schome Park environment, and those that engaged further in related Wiki and forum discussions enhanced their knowledge age skills (e.g., communication, teamwork, leadership, and creativity). Further, students reported a sense of safety in the Schome Park environment, a result which was particularly evident amongst participants from socially disadvantage groups or ethnic minority backgrounds. One disadvantage found in the schome-

NAGTY pilot related to the substantial learning curve that students undertook initially in order to navigate and interact within the Second Life environment.

Although to date, the use of Second Life in specific relation to assessment has been at a discussion level, with no actual curriculum assessments being currently available via this media, the resources available within this virtual environment.

Educational Outcomes

Assessment Personalisation – Personalised Learning and Assessment Spaces

Albeit in relation to the earliest and by today's standard crude PC technology, Turkle (1984) recognised the value of computers as being "the second self," for learning situations where the student's personality, age, and style of learning requires a personalised approach. For clarity, it should be noted that *personalised* learning is different from *individualised* learning. Individualised learning represents students working individually on tasks generated from a pre-determined educational program, typically students are left to their own devices, resulting often in low learning and performance aspirations (DECS Youth Engagement Team, 2005). Personalised learning refers to a highly structured and responsive approach to a student's learning and assessment. Students are typically provided with choices within their learning objectives and curriculum, and presents a context where the student's learning is by there own passions and interests (DfES, 2006). In a world where there were only a few students to a class, the personalisation of each student's optimal learning needs and learning style would be prescribed and managed by the teacher (Poole, Sky-McIlvain, & Jackson, 2006). This of course, is not feasible when class numbers are large (i.e., low teacher-student ratio), and where typically there are only a limited amount of learning resources available. In such cases, many recent technologies can increase the degree of personalisation.

Whilst, technology does allow a student's assessment to be integrated and managed across individual student's needs, with such solutions providing extensive progress tracking and target setting relating to individual students, research has indicated that it is rare for this information to progress from the teacher to the student. Ofsted (2005) argued that a specific emphasis should be given to getting performance information to students in a form that is intelligible and specific to their own learning status and requirements. Students need to have a feedback mechanism that makes available their performance, tracks their progress, and shows their progress in relation to set target levels (Ofsted, 2005). These feedback issues relate to the starting point for the delivery of personalised learning, namely *assessment for learning*. This component relates to the learning evidence and assessment feedback supplied to the student so that they can identify where they are at in their learning, what they are required to do next, and how this 'what next' aspect is achieved.

One approach for integrating student personalised learning and assessment that has gained momentum over the past five years, are e-portfolios. E-Portfolios refer to the personalised spaces that enable learners to save their own work, store material that as been or as yet to be assessed, maintain records of their assessment and achievements, and allow access to personal records and work (Twining et al., 2006). As a personalised mechanism, e-portfolios represent a move towards a self-directed, learner-centered approach to learning that incorporates both autonomous learning and peer-to-peer (Roberts et al. 2005). Thus, student own and manage their own personal bank or repository of learning and assessment evidence, experiences, information, and feedback. Research on the adoption of e-portfolios have found that in addition to students becoming more self-directed in their learning approach, reflective learning and practice have also been developed (e.g., Cotterill, Bradley & Hammond, 2006; Driessen, Van Tartwijk, Overeem, Vermunt, & Van der Vleuten, 2005). Thus, student's learning ownership of their learning appears to impact on their motivation to learn in the classroom. An Evidence for Policy and Practice Information and Co-ordinating Centre (Eppi Centre) Review (2005) study of what factors impact on students' motivation to learn in the classroom found that student are more likely to be engaged with their curriculum if they believe it is relevant and if they are given the opportunity to have ownership of their learning and reflective of their achievement. A major limitation to many uses of e-portfolios is that they are poorly constructed and the evidence included is rarely considered from an assessment viewpoint. The typical advice to ameliorate this problem is for the teacher (and student) to agree on a scoring rubric prior to designed the e-portfolio to thence maximise the feedback information, ensure they are growing and not static artefacts, and to ensure that there is progression to desired targets (Johnson et al., 2008).

Regarding the delivery of assessments via the e-portfolio platform, two studies have investigated the effectiveness and efficiency of applying traditional testing approaches to this environment. Oosterheert, van Eldik, & Kral (2007) investigated summative competency assessment across through different models. They present that argument that an inflexible, prescribed curriculum context restricts the value produced from e-portfolios. In addition, Oosterheert et al. argued that e-portfolios holistic learning platform must result in a trade-off between holistic competency assessment and 'isolated' tests. Further, they posited that the use of many current methods of standardised assessment, whilst increasing the reliability, does an injustice to the authenticity of students learning and learning contexts. Thus, the problematic proposition arises of what should be standardised in e-portfolio assessments. According to Driessen et al. (2005) and Baume and York (2005), assessment criteria should be standardised, as this will increase the assessors' inter-subjectivity and thus the reliability of their assessments. Although Van der Vleuten and Schuwirth (2005) agreed that the assessment process should be standardised they argued that the objective should be that the test-takers reach a "saturation point" whereby they approve of their assessments (e.g., face validity). As a result, Van der Vleuten and Schuwirth proposed that more the subjective, unstructured and non-standardised assessments will then become more reliable.

Assessment of classroom behaviours

Educators, teachers and educational psychologists have long studied classroom environments, and this research has significantly impacted on the understandings of optimal instructional practice, the training and development of teachers and curriculum (Hinely & Ponder, 1979). However, in comparison to cognitive assessment, investment in classroom measurement has been negligible. Nuthall (2000) once argued that "the central problem of research on teaching is still the problem of creating an evidence-based model of the learning process that can be used by teachers to decide how classroom activities are affecting the learning process of individual students" (p. 4). To date, the issue outlined by Nuthall still exists. Building from Nuthall and Alton-Lee's work, this section will focus on the current status of the assessment of classrooms, drawing technological advancements, and outline options for future classroom developments and observations.

Status of current assessment of classrooms

A high-quality classroom environment is an important mechanism for ensuring positive outcomes for students (Grinder, 2007). Thus, it is vital not only to gather information on a student's assessment outcomes, but also in relation to the quality of the classroom environment, particularly, as children's developmental outcomes tend to be dependent on the quality of their experiences in educational (and family) settings (Pianta, 2003). Traditionally, the assessment of classroom quality has been conducted via the use of standardized classroom observation systems. This measurement approach is typically been directed towards assessing a multitude aspects of classroom life, providing valuable information on teachers' strengths and weaknesses, professional development, evaluating of programs/interventions, and evaluating the impact of policy initiatives (Dorman, Aldridge, & Fraser, 2006).

Reviews on research relating to classroom environment (e.g., Dorman, 2002; Fraser, 1998; Goh & Khine, 2002) have established ten distinct areas where research has been directed:

1. Associations between classroom environment and outcomes
2. Evaluation of learning interventions or innovations
3. Student and teacher differences between perceived classroom environment/experiences
4. Comparisons between actual and preferred classroom environments

5. Impact of variables (e.g., gender, year level, school type, subject) on classroom environment
6. Transition from primary to secondary school environments
7. School psychology
8. Teacher education
9. Teacher/student productivity
10. Impact of classroom resources and tools on classroom life

Thus, results from the areas across these studies have confirmed the powerful impact that the *quality* of the classroom environment has on student outcomes (Dorman, Alridge, & Fraser, 2006). For example, students who participate in high-quality settings score higher on standardized measures of 'school readiness' and are engaged in more complex activities with their peers (Burchinal, Lee, & Ramey, 1998; Campbell & Ramey, 1994; Howes, & Hamilton, 1993; Kontos, Howes, Shinn, & Galinsky, 1997). Even research focusing on preschoolers who have attended high-quality learning environments have found that they typically enter Kindergarten with a greater knowledge of verbal and numerical concepts, increased ability to cope with school tasks, higher likelihood of attaining at least normal progress through their primary grades, and display higher ratings of interactive and social skills (e.g., La Paro, Pianta, & Stuhlman, 2004).

<u>Future options for classroom observations</u>

Hand-held technology provides an invaluable tool for the rapid collection of data from focus areas of teaching learning. The University of Kentucky utilised hand-held computers for a project where the principal completed conducted five classroom walkthrough observation studies to assess the learning and teacher indicators (e.g., teacher and student engagement, types of classroom activities, lesson planning and instructional differentiation, and grouping arrangements) displayed by within the classroom environment. Many studies have shown that so many classrooms follow the "recitation script" (Tharp & Gallimore, 1988). An *initiation*, usually in the form of a teacher question, a *response* in which the student aims to answer in the question, and a *follow-up* move, in which the teacher provides some form of response. This is prevalent in teacher-directed sessions, and in so many peer sessions the three-part exchange is evident among peers. If students are to be asked to have more role in their own learning, if teachers are to better understand how to use a student's prior knowledge and current understandings, if peer and student talk is so powerful in the learning equation, then a greater understanding of how students engage in their learning, with their teachers and peers, is of considerable value (Nuthall, 2000).

Hardman and his colleagues (see Smith & Hardman, 2003) have used technology to assist teachers to better understand their classrooms. His Computerised Interaction System (CIS) works via a hand-held device about the size of a calculator. It then enables observation and recording of the lesson in real-time and the results are available for immediate analysis. Compared to the older pencil-and-paper systems it is quicker, mobile and immediate, it is possible to retrospectively re-analyse classrooms from the videos captured, it is highly adaptable (coding can change on the run), it is relatively non-intrusive, and the coding is highly reliable. Ackers and Hardman (2001) showed the dominance in most classrooms of the transmission mode of teaching and the three-part exchange that tends to dominant interactions (teacher question, student response, teacher reaction); a prescriptive literacy strategy leads to more not less prescriptive teaching especially when the strategy demands that there be greater student participation and involvement (Hardman, Smith, & Wall, 2003); and effective teachers have greater frequency but similar types of interactive discourse (Har

However, studies of classroom discourse from North America and the United Kingdom (e.g. Mehan, 1979; Dillon, 1994; Edwards & Westgate, 1994) show that whole class teaching across all stages of schooling is dominated by what Tharp & Gallimore (1988) call the 'recitation script'. In its prototypical form teacher-led recitation consists of three moves: an *initiation*, usually in the form of a teacher question, a *response* in which a student attempts to answer the question and a *follow-up* move, in which the teacher provides some form of feedback (very often in the form of an evaluation) to the pupil's response. This

huma
incid
syste
At th
pion

Software Applications

Assessment Scoring

Automated scoring refers to the score that is produced automatically or in real-time on writing that is constructed, e.g., either short-answer or essay questions. The scoring method adopted for assessments is associated with both the item format and response action. Historically, selected-response items, such as a multiple-choice item, simply required the computer algorithm to capture the responses from test-takers, and then automatically and simultaneously capture a score. However, the inclusion of innovative items that have various constructed responses formats, present a distinct challenge and potential constraint to the use of these designs (Zenisky & Sireci, 2002). This section examines the software that has been developed specific relation to the scoring of responses from test-takers constructed responses, across three distinct scoring approaches as classified by Zenisky and Sireci - automated scoring of essays, expert systems analysis, and mental modeling. The following outlines some of the most recent and innovative software available and analysis approaches under each of these classifications (for earlier technologies, refer to Zenisky & Sireci's review).

<u>Automated Scoring</u>

Whilst essay writing and short answers are the most common type of constructed response type found in assessment, there use has been, until recently, restricted due to the requirement for at least two subject-matter experts to evaluate each written response. Given the inherit time, resource costs, reliability, and generalisability issues associated with written items, it is no surprise that there has been significant attempts to try and totally automate the scoring requirement. Another encouraging feature of this computer technology is the overwhelming high agreement rates that has been established between automated scoring systems and subject-matter experts (e.g., Attali, 2004; Burstein & Chodorow, 1999; Landauer, Laham, & Foltz, 2003; Nichols, 2004; Page, 2003).

Educational Testing Service (ETS) has produced four products that utilised the latest developments in natural language processing technologies in order to evaluate and score the writing produced by test-takers. Specifically, the natural language processing techniques identify the syntactical and lexical cues in the writing. The automated scoring program, Criterion, that allows the teacher to select from over a hundred writing tasks (called prompts) which span all grade levels, across narrative, expository, and persuasive curriculum domains. Each prompt has been extensively developed to be calibrated against either a four- or six-point scoring rubric, and offers a range of representative essays (ETS Technologies Inc, 2002). Criterion's web-based automatic scoring engine utilised two other innovations, e-rater and Critique to produce immediate real-time scoring, within seconds of essay's online submission. E-Rater is a scoring engine which produces scores for holistic responses, whilst Critique provides real-time feedback reporting on grammar, punctuation, and spelling, in addition to the mechanics, style, and organisation of the writing. Another technological innovation is the off-topic feature of Criterion, specifically, the program has been developed to identify when the essay submitted is not actually answering the essay topic. Previously, automated essay scoring solutions have been criticised for their vulnerability to cheating (Kukich, 2000; Rudner & Gagne, 2001). This feature is powerful in relation to the attempts of some students to use commercially acquired essay which are simply 'cut-and-pasted' as the response (Attali, 2004).

Another dominant automatic scoring system is IntelliMetric, and its student feedback system MY Access! created and developed by Vantage Learning. Using artificial intelligence, natural language processing, and statistical technologies, IntelliMetric is a classed as a 'learning engine' that is described as internalising the collective wisdom of subject-matter expert raters (Elliot, 2003). As an advanced AI application for scoring essays, IntelliMetric relies on Vantage Learning's CogniSearch and Quantum Reasoning technologies (Vantage Learning, 2001a, 2003c). Using this blend of technologies, Vantage Learning state that the scoring system's artificial intelligence allows the software to 'learn' the writing response characteristics that human raters would apply high and conversely, low value to (Shermis & Barrera, 2002; Shermis et al., 2003). An added feature of the IntelliMetric approach is it's capability to evaluate essay responses across multiple languages

– English, Spanish, Hebrew, Bahasa, Dutch, French, Portuguese, German, Italian, Arabic, and Japanese (Elliot, 2003). As mentioned above, the score obtained form the IntelliMetric software is integrated with the MY Access! feedback system. This software provides immediate scoring and diagnostic feedback on the essay that they have submitted, and supplies writing tools (writing dashboard and a writing e-portfolio) to allow the student to see their own progress, list their completed assignments, feedback reports, and various diagnostic details.

Conclusion

Increasingly technology as discussed in this paper is being used as a catalyst to help enrich the assessment learning environment that is provided to students. Although less research has been conducted in relation to applying these technologies to the assessment realm, it is only a matter of time before the dynamic teaching learning environments created by the technologies are extended to the assessment arena. However, the use and promise of technology for learning and assessing that learning is reliant on a paradigm shift from traditional teaching techniques to a dynamic and interactive paradigm (Glenn & Melmed, 1995). It is not obvious which will come first, but it is certain that students are more likely to embrace the newer technologies and thus be major agents for innovation.

Education is moving towards providing students with more personal learning opportunities, and for teachers to see learning more through the "eyes of the students". Individual ownership of learning and progress are proposed as being essential requirement of learners and are achievable through e-portfolios. Allowing students the portability and flexibility of mobile hand-held devices also enhances the access and engagement offered by material and feedback delivered through these mechanisms. However, while developments using technology to personalise assessment whether via software or hardware devices are evident, they are still in there infancy (Becta, 2007).

Computer technology is become more invisible – inherent in many daily devices (telephones, microwaves, PDAs, televisions, CCTV) although when computers are referred to in school contexts there is still a vision of a box and screen of some sort. The speed and storage now available will allow many innovations in multimedia and computer adaptive technologies, and the recent innovations in virtual reality and second life can be used in more efficient and effective assessment. At the moment, where these technologies are being used their success is more measured in engagement than information gained. Only if the assessments are more dependable (valid and reliable) as well as more efficient (time, intrusiveness, immediacy, personalised) is it likely that these methods will replace many of the sturdy and well tried and known (usually pencil-and-paper) methods. Further, the use of innovative and valid Reports are likely to drive this change rather than the technology itself. It is anticipated that the greatest gains from introducing many of these assessment technologies will be shift the locus of classrooms more from the ‘teacher to the student’ towards the ‘student to and with the teacher’.

This paper has presented some of the applications that have been developed from evolving technologies for the learning and assessment of knowledge. Whilst presented separately in this paper, these hardware and software technologies can act together as being both the mode from which learning takes place and the medium for assessing the acquired knowledge. These two separate functions are inherently linked, specifically, the method of assessment must be a reflection of the technological tools that were adopted to present the teaching and learning (CEO Forum on Education and Technology, 2001 as cited in Bennett, 2002). Thus, there should not exist a mismatch between these two functions of technology. For example, research by Russell and Plati (2001) showed that student performance on a writing proficiency via a paper-and-pencil test underestimated their proficiency, given the level of writing composition displayed when completing assignments onscreen. As Bennett (2002) stated, asking students to convey their learning in a medium that is different from the one they routinely use will become an untenable approach.

REFERENCES

Ashworth, S. D., & McHenry, J. J. (1992, April). *Developing a multimedia in-basket: Lessons learned.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Attali, Y. (2004). *Exploring the feedback and revision features of Criterion.* Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA.

Becta (2007). *Harnessing technology review 2007: Progress and impact of technology in education.* Coventry: Becta.

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*, from http://www.jtla.org

Breland, H. M. (1999). Exploration of an automated editing task as a GRE writing measure. (GRE Board Professional Report No. 96-01R). Princeton, NJ: Educational Testing Service.

Burchinal, M., Lee, M., & Ramey, C.T. (1998). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development, 60*, 128-137.

Burchinal, M., Lee, M., & Ramey, C.T. (1998). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development, 60*, 128-137.

Burstein, J., & Chodorow, M. (1999, June). Automated essay scoring for nonnative English speakers (PDF). In *Proceedings of the ACL99 workshop on computer-mediated language assessment and evaluation of natural language processing*. College Park, MD.

Campbell, F.A., & Ramey, C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low income families. Child Development, 65), 684-698.

Campbell, F.A., & Ramey, C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low income families. *Child Development, 65*, 684-698.

Cotterill, S., Bradley, P., & Hammond, G. (2006). ePortfolios: supporting assessment in complex educational environments. In Bryan, C. & Clegg, K. (Eds.). *Innovative assessment in higher education*. London: Routledge, 191-199.

DECS Youth Engagement Team Report. (2005). *Engaging young people in remote area schools.* Government of South Australia.

Desmarais, L. B., Masi, D. L., Olson, M. J., Barbera, K. M., & Dyer, P. J. (1994, April). *Scoring a multimedia situational judgment test: IBM's experience.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Desmarais, L. B., Dyer, P. J., Midkiff, K. R., Barbera, K. M., Curtis, J. R., Esrig, F. H., & Masi, D. L. (1992, May). *Scientific uncertainties in the development of a multimedia test: Trade-offs and decisions.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Montreal Quebec.

DfES. (2006). *National curriculum assessments at Key Stage 2 in England* from http://www.dfes.gov.uk/rsgateway/DB/SFR/s000697/SFR48-2006.pdf

Dorman, J. P. (2002). Classroom environment research: Progress and possibilities. *Queensland Journal of Educational Research, 18,* 112-140.

Dorman, J. P., Aldridge, J. M., & Fraser, B. J. (2006). Using students' assessment of classroom environment to develop a typology of secondary school classrooms. *International Education Journal, 7,* 906-915.

Drasgow, F., & Mattern, K. (2006). New tests and new items: Opportunities and issues. In D. Bartram & R. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances*. London: Wiley.

Drasgow, F., Olson-Buchanan, J. B., & Moberg, P. J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 177-196). Mahwah, NJ: Lawrence Erlbaum Associates.

Driessen, E.W., Van Tartwijk, J., Overeem, K., Vermunt, J.D., & Van der Vleuten, C.P.M. (2005b). Conditions for successful reflective use of portfolios in undergraduate medical education. *Medical Education, 39*, 1230-1235.

Dyer, P. J., Desmarais, L. B., Midkiff, K. R., Colihan, J. P., & Olson, J. B. (1992, May) *Designing a multimedia test: Understanding the organizational charge, building team, and making the basic research commitments.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec.

Elliot, S. (2003). IntelliMetric: from here to validity. In Mark D. Shermis and Jill C. Burstein (Eds.). *Automated essay scoring: a cross disciplinary approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Eppi Centre Review. (2005). *A systematic review of what pupils, aged 11-16, believe impacts on their motivation to learn in the classroom.* Retrieved June 26, 2008 from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid+304

ETS Technologies Inc. *Criterion: Online writing evaluation.* Princeton: ETS.

Fraser, B. J. (1998). Classroom environment instruments: Development, validity, and applications. *Learning Environments Research, 1,* 7-33.

Glennan, T. K., & Melmed, A. (1995). *Fostering the use of educational technology: Elements of a national strategy.* Washington, DC: RAND, from http://www.rand.org/publications/MR/MR682/

Glennan, T. K., & Melmed, A. (1995). *Fostering the use of educational technology: Elements of a national strategy.* Washington, DC : RAND . Retrieved September 21, 2005, from http://www.rand.org/publications/MR/MR682/

Goh, S. C. & Khine, M. S. (2002). *Studies in educational learning environments: An international perspective.* Singapore: World Scientific.

Grinder, E. (2007). *Review of early childhood classroom observation measures.* Harrisburg, PA: PA Department of Education.

Hinely R., & Ponder, G. (1979). *Theory, practice, and classroom research.* Lawrence Erlbaum Associates (Taylor & Francis Group): United Kingdom.

Howes, C., & Hamilton, C.E. (1993). Child care for young children. In B. Spodek (Ed.) *Handbook of research on the education of young children* (pp. 322-326). New York: Macmillan.

Howes, C., & Hamilton, C.E. (1993). Child care for young children. In B. Spodek (Ed.) *Handbook of research on the education of young children* (pp. 322-326). New York: Macmillan.

Keenan, P. A., & Olson, J. B. (1991, April). *A model-based multi-media technique for assessing conflict management skills.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology Conference, St. Louis, MO.

Kontos, S., Howes, C., Shinn, M., & Galinsky, E., (1997). Children's experiences in family child care and relative care as a function of family income and ethnicity. *Merrill-Palmer Quarterly, 43,* 386-403.

Kontos, S., Howes, C., Shinn, M., & Galinsky, E., (1997). Children's experiences in family child care and relative care as a function of family income and ethnicity. *Merrill-Palmer Quarterly, 43,* 386-403.

Kukich, K. (2000, September/October). Beyond Automated Essay Scoring. In M. A. Hearst (Ed.), *The debate on automated essay grading*. IEEE Intelligent systems, 27–31. Retrieved November 12, 2004, from http://que.infoscience.uiowa.edu/~ light/research/mypapers/autoGradingIEEE.pdf

La Paro, K. M., Pianta, R.C., & Stuhlman, M. (2004). The Classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal, 104 (5)*, pp. 409-426.

Lagorio, C. (2007). The ultimate distance learning [Electronic version]. *Journal of Bibliographic Research, 5*, from http://www.nytimes.com/2007/01/07/education/edlife/07innovation.html?_r=1&oref=slogin

Landauer, T., Laham, D. & Foltz, P. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor™. In M. Shermis & J. Burstein (Eds.), *Automated essay scoring: A crossdisciplinary perspective* (pp 87–112). New Jersey, USA: Lawrence Erlbaum Associates.

Leeson, H. V. (2008). *Maxmizing information: Applications of ideal point modeling and innovative item design to personality measurement.* Unpublished manuscript, University of Auckland.

Maddison, S., & Lorincz, G. (2003). Bridging the digital divide. *Computing & control eng*

Martin, W., Heinze, C. & Kanaya, T. (2005). Trends in use of handheld technology to support student reading assessment. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology and Teacher Education International Conference 2005* (pp. 134-137). Chesapeake, VA: AACE.

McHenry, J. J., & Schmitt, N. (1994). Multimedia testing. In M. G. Rumsey & C. B. Walker (Eds.), *Personnel selection and classification* (pp. 193-232). Hillsdale, NJ: Lawrence Erlbaum Associates.

Midkiff, K. R., Dyer, P. J., Desmarais, L. B., Rogg, K., & McCusker, C. R. (1992, May). *The multimedia test: Friend or foe?* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology Conference, Montreal, Quebec.

Nuthall, G. (2000). Understanding how classroom experience shapes students minds. *Paper presented at the 17, Kongress der Deutschen Gesellschaft fur Erziehungswissenschaf,* Gottingen, Germany.

Nuthall, G., & Alton-Lee, A. (1997). *Student learning in the classroom: Report to the Ministry of Education.* Christchurch: University of Canterbury.

Ofsted (2005). *Embedding ICT in schools – a dual evaluation exercise.* London: Ofsted.

Oosterheert, I., van Eldik, S., & Kral, M. (2007). The ePortfolio as an instrument for summative competency assessment. In A. Wijnand & M. Veugelers

(Eds.), *Stimulating lifelong learning: The ePortfolio in Dutch higher education* (pp. 55-61). SURF NL: Portfolio.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, *51*, 1-24.

Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). Mahwah, NJ: Lawrence Erlbaum Associates.

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice*, (pp. 129-148). Dordrecht, The Netherlands: Kluwer.

Pianta, R.C. (2003). Standardized classroom observations from pre-K to third grade: A mechanism for improving quality classroom experiences during the P-3 years from www.ffcd.org/news/publications.html

Renaudon-Smith, J. (2008). *Technology enabled education a catalyst for positive change.* [PowerPoint] from http://209.85.173.104/search?q=cache:KCWExJFwqbgJ:linc.mit.edu/conference/presentations/RenaudanSmith.ppt+ink+media+technology+enabled+education&hl=en&ct=clnk&cd=1&gl=nz

Roberts, G., Aalderink, W., Cook, J., Feijen, M., Harvey, J., Lee, S., & Wade, V. P. (2005, April). *Reflective learning, future thinking: Digital repositories, e-portfolios, informal learning and ubiquitous computing.* Paper presented at the ALT/SURF/ILTA Spring Conference Research Seminar. Retrieved June 23, 2008, from http://www.alt.ac.uk/docs/ ALT_SURF_ILTA_white_paper_2005.pdf

Roschelle, J., Pea, P., Hoadley, C., Gordin, D. & Means, B. (2000) Changing How and What Children Learn in School with Computer-Based Technologies. *The Future of Children Children and Computer Technology*, 10 (2). Retrieved September 8, 2004, from http://www.futureofchildren.org/usr_doc/vol10no2Art4.pdf

Rudner, L. & Gagne, P. (2001). *An overview of three approaches to scoring written essays by computer* (ERIC Digest number ED 458 290).

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *TC Record.Org.* Retrieved July 27, 2008 from http://www.tcrecord.org/Content.asp

Shermis, M. & Barrera, F. (2002). *Exit assessments: Evaluating writing ability through Automated Essay Scoring* (ERIC document reproduction service no ED 464 950).

Shermis, M. D., Raymat, M. V., & Barrera, F. (2003). *Assessing writing through the curriculum with Automated Essay Scoring* (ERIC document reproduction service no ED 477 929).

Somekh, B., Underwood, J., Convery, A., Dillon, G., Jarvis, J., Lewin, C., Mavers, D., Saxon, D., Sing, S., Steadman, S., Twining, P., & Woodrow, D. (2007). *Evaluation of the ICT Test Bed project final report.* Coventry: Becta.

Trinder, J. J., Magill, J. V., & Roy, S. (2005). Portable assessment: Towards ubiquitous education. *International Journal of Electrical Engineering Education, January* from http://www.manchesteruniversitypress.co.uk/information_areas/journals/ijeee/IJEEE.htm

Twining, P., Broadie, R., Cook, D., Ford, K., Morris, D., Twiner, A., & Underwood, J. (2006). *Educational change and ICT: An exploration of priorities 2* *and 3 of the DfES e-strategy in schools and colleges.* Coventry: Becta.

Vantage Learning. (2001a). *About IntelliMetric* (PB-540). Newtown, PA: Vantage Learning.

Vantage Learning. (2003c). *A true score study of 11th grade student writing responses using IntelliMetric Version 9.0* (RB-786). Newtown, PA: Vantage Learning.

Zenisky, A.L., & Sireci, S.G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15*, 337-362.