

Challenges for Information Technology supporting Educational Assessment

Mary Webb, *mary.webb@kcl.ac.uk*,

King's College London, UK

David Gibson, *david.gibson@curveshift.com*,

Curveshift, USA

Alona Forkosh-Baruch, *alonabar@post.tau.ac.il*

Tel-Aviv University, Israel

Keywords

Assessment, computerised assessments, personalisation, formative assessment, quiet assessment, summative assessment, assessment FOR learning, stealth assessment

ABSTRACT

This paper examines the scope for IT-enabled assessments to serve simultaneously both learners and the enterprise of education. The paper has its origins in global discussions initiated with the publication of the International Handbook of Information Technology in Primary and Secondary Education and continued in subsequent international EdusummiT conferences. These discussions identified a need for alternative assessment approaches designed on new understandings of the impact of ICT on learning, assessment and society. In order to move this debate forward this paper examines theoretical frameworks that underpin assessment and can guide assessment system design as well as implementation processes. The paper proposes ways of combining frameworks that come from two different perspectives: 1) a conceptual approach to assessment design for computerised assessment based on evidence centred design (ECD) and 2) a framework for formative assessment (assessment FOR learning) based on empirical research in classrooms. The paper argues that combining the ECD and formative assessment frameworks and building on the opportunities provided by computerised assessments as well as harnessing teachers' and students' experience and developing their validation processes could enable assessments to address simultaneously assessment FOR learning and assessment OF learning. Strategies would include harnessing the benefits of embedded continuous unobtrusive measuring of performance while learners are engaged in interesting computerised tasks that have been designed to support their learning. Learners need to be involved in discussing and negotiating their learning so we conceptualise these embedded unobtrusive processes as "quiet assessment", whose volume can be turned up by learners whenever they wish, to give them access to meaningful representations of arguments about their achievements. These strategies could enable a wider range of measures to contribute to judgements of students' achievements thus supporting their learning in 21st Century contexts.

INTRODUCTION

This paper examines the scope for IT-enabled assessments to serve simultaneously both learners and the enterprise of education. The paper has its origins in a global discussion initiated with the publication of the International Handbook of Information Technology in Primary and Secondary Education (Voogt & Knezek, 2008) and continued in subsequent discussions of the Assessment Working Group at EdusumMIT 2011 (Webb & Gibson, 2011) subsequently summarised in a final report (Resta, Searson, Patru, Knezek, & Voogt, 2012). Three aspects of assessment identified as critical for 21st century assessment models were: student involvement in assessment, digitally-enhanced assessment, and assessment of application of ICT skills acquired in formal and informal learning environments. Digitally-enhanced assessments were defined by the Working Group as those that integrate 1) an authentic learning experience involving digital media with 2) embedded continuous unobtrusive measures of performance, learning and knowledge, e.g. “stealth assessment” (Shute, 2011) which 3) creates a highly detailed (high resolution) data record which can be computationally analyzed and displayed so that 4) learners and teachers can immediately utilize the information to improve learning. The Working Group recommended that education stakeholders should articulate and support a shift in assessment from the dominant high stakes, test-based evaluation model used in many school systems, to [a balance of models](#) that more effectively measure the preparedness of today’s digital age learners.

In order to measure the complex, higher order outcomes empowered by technology-enriched learning experiences, there is a need for alternative assessment approaches and instruments that surpass the limitations of current assessment systems and approaches and are designed on new understandings of the impact of ICT on learning, assessment and society at large (Erstad, 2008). New technologies are expanding the range of possibilities for assessments, including increasing opportunities for personalisation of assessments (Yeh, 2010) and the

1/25/13 3:31 PM

Deleted: one

1/25/13 3:32 PM

Deleted: s

capability for assessment to measure a broader range of knowledge and knowledge-in-action (Clarke & Dede, 2010). For example students can be assessed through simulations, e-portfolios and interactive games (Gibson, Aldrich, & Prensky, 2007). However assessment exists in a complex dynamic relationship with curriculum, pedagogy, and the needs and demands of learners and of the world outside of schools (Forkosh-Baruch, Gibson, Schulz-Zander, & Webb, 2009). For example, innovative developments in formative assessment can support and even drive needed developments in pedagogy (Black, Harrison, Lee, Marshall, & Wiliam, 2003). At the same time assessment systems have become increasingly complex as they are expected to serve a wide range of purposes. Assessments are not only about supporting learning but their purposes also range from judging individual students, to evaluating schools and districts and monitoring national performance (Klenowski & Wyatt-Smith, 2011; Mansell, James, & Group, 2009). Some of these purposes are in tension with each other (Mansell, et al., 2009). For example, current high stakes assessments tend to constrict the curriculum and to discourage the development of creativity and the social and cultural aspects of learning (Harlen and Deakin Crick 2002). Therefore designs for new assessment systems using new technologies need to focus on finding ways to assess while also encouraging the development of these higher order, more complex skills.

The future of assessment in the digital age may appear in many forms. It may involve for example a pedagogical agent patiently tutoring someone in anything he or she would like to learn (Sabourin, Mott, & Lester, 2011); an analysis of a learner's decisions during a digital game or simulation (Clarke & Dede, 2010; Gibson, 2011); students reviewing and commenting on each others' digital creations through an online discussion (Ertmer et al., 2007; Van Der Pol, Van Den Berg, Admiraal, & Simons, 2008; Webb, 2010); a multimedia-constructed response item created with an online animation and modeling application (Lenhard, Baier, Hoffmann, & Schneider, 2007; Mislevy, Steinberg, & Almond, 2003); students receiving remote asynchronous

1/25/13 3:38 PM

Formatted: Not Highlight

1/25/13 3:38 PM

Formatted: Not Highlight

expert feedback about how they worked with each other via ICT to solve a problem and communicate their understandings (Rissanen et al., 2008) or an emotionally engaging virtual world experience that unobtrusively documents progression of a person's leadership and ethical development over time (Turkay & Tirthali, 2010). Whether these possibilities excite or concern us, this small set of vignettes begins to outline a broad range of possibilities that place ICT in a variety of roles including a medium for communication, learning assistant, judge, test item and performance prompt, practice arena, and performance workspace.

In this paper we will examine the role of ICT in assessment in relation to: 1) the nature of assessment, by which we mean the essential qualities or characteristics by which assessment is recognized; and 2) the potential of technological advances for transforming assessment processes. In this endeavour, we will focus particularly on the possibilities and challenges for ICT-enabled assessment serving simultaneously the needs of the learner and the enterprise of education.

THE NATURE AND PURPOSES OF ASSESSMENT

Phrases in common use in education, such as assessment *OF* learning, assessment *FOR* learning and assessment *AS* learning illustrate the complexity of defining assessment (Bennett, 2010). The limitations of the traditional distinctions between formative assessment and summative assessment have been discussed extensively ((Black & Wiliam, 2009; Harlen & James, 1997). Attempts have been made to clarify assessment for teachers by using the term "assessment for learning" in place of formative assessment (see Wiliam, 2011 for an explanation of the origin of this term). Assessment for learning was popularised in the UK by Black et al (2002) and in the US by Stiggins (2005). Our working group discussions suggested that the term assessment for learning is now used widely but recent experience in the UK suggests that confusion remains: for example the UK government introduced a frequent summative testing regime (Assessing Pupils' Progress) on the basis of evidence for the

effectiveness of assessment for learning (Black, Harrison, Hodgen, Marshall, & Serret, 2010 P. 216)

The nature of assessment thus requires some clarification because the word is used for both a process of acquiring evidence and a label for a judgment of results based on specific worthiness. We need to know whether the purpose and effect of a particular assessment is educative (Wiggins 1999) / formative (FOR learning) or is it to assess the extent of an attained state of education (OF learning), or whether both purposes can be achieved through the same process. In addition, the word “assessment” is used for both the act of judging the evidence and the results of that act. These considerations and use of terms can lead to confusion in conversations, policy circles, and in practice (Harlen & James, 1997), because there are at least four ways to think about assessment, all of which can be supported, enabled or enhanced by ICTs (Table 1).

One perspective is of the formative / educative process – learning from the feedback information provided by an assessment process. Another perspective focuses on formative results for instructional adaptation – the decision to make improvements based on an interpretation of evidence from assessment. The third is the understanding of the assessment and the degree of engagement by students and teachers in the act of assessment. And the fourth is the value judgment of results – that provides summative information based on evidence.

1/25/13 3:39 PM
Formatted: Not Highlight

	PROCESS focus	RESULTS focus
Assessment FOR learning	1) Feedback information	2) Improvement Decisions
Assessment OF learning	3) Degree of engagement with / understanding of process	4) Value Judgments

Table 1. Four ways to think about assessment

Assessment is said to be educative when it produces information that is used to create and support learning opportunities, while a summative assessment is more judgemental, focusing on results by determining the validity and value of performances and attainments (Wiggins, 1999).

Black and William further refined their earlier use of the term formative to focus on the importance of decisions:

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & William, 2009 p.9)

Thus educative and formative are used in relation to assessment practice in similar ways in these definitions. Perspectives 1, 2 and 3 in Table 1 are important for assessment to be formative / educative while perspectives 3 and 4 are essential for summative assessment. Perspective 3 is important for both formative and summative assessment because in order to generate valid assessment information students need to engage with and understand the process and its purpose. Perspective 4 is about making summative judgements for purposes of grading, accreditation, setting and other more general evaluative purposes. While such judgements are important and necessary at particular transition points they may hinder formative / educative processes if used too often or for too many diverse purposes. For example, evidence suggests that students fail to attend to feedback comments when given grades (Butler, 1988). Furthermore using assessment evidence for a range of different purposes may result in complex effects on validity as discussed later.

CONCEPTUALISING ASSESSMENT DESIGN

An important challenge for the design of educational assessment is whether all four perspectives on assessment: feedback information, improvement decisions, degree of engagement and understanding, and value judgments can co-exist to the benefit of learners. In

order to explore this challenge in more depth we will examine and compare two important current frameworks for assessment design: 1) Evidence-Centered Design (ECD) for computerised assessments and 2) a framework for formative assessment (assessment FOR learning) based on empirical research in classrooms. The ECD framework (Mislevy, et al., 2003) was developed by members of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) in the USA and has been applied to the design of a range of different types of computerized assessments including adaptive assessments, games, simulations (Rupp, Gushta, Mislevy, & Shaffer, 2010; Shute, 2011) and portfolios (Gibson, 2010). ECD is a four-staged assessment development framework that includes domain analysis; domain modelling; operational processes and the delivery system. The ECD framework makes explicit the interrelations among substantive arguments concerning domain models, validity, assessment designs, and operational processes (Mislevy, et al., 2003). The framework has diagnostic capabilities and provides for opportunities for stakeholders to view estimated competency levels, examine the evidence on which these judgments were based and to use this information for educative, formative purposes if appropriate (Shute, 2011 P. 9).

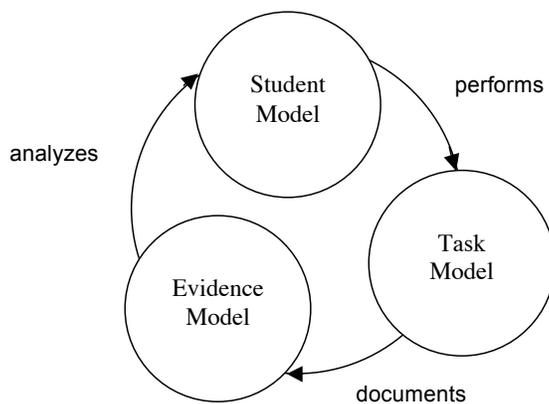


Figure 1. The conceptual assessment framework.

The operational processes in the ECD include three submodels (Figure 1); the student, task and evidence models, known as the Conceptual Assessment Framework (CAF) within the ECD. Each of these three components are flexible, adaptive concepts that evolve within the computerized assessment context. For example, if knowledge in the domain changes, then the task model changes; as a student attempts a task, assistance and information can be used to scaffold task performance and adapt the task; and evidence models adapt and change as the student and task models change. The student model captures the idea that our conceptions of good performance and satisfactory evidence of domain knowledge guide, and constrain, what a student does to demonstrate that evidence. The task model defines a required performance; and the evidence model determines the interpretation method and the rules and criteria for making assessment determinations from the performance data.

To help ensure that the designs of computerised assessments take account of learners' needs, we next examine and compare a framework derived from empirical work on formative assessment in classrooms (Black & Wiliam, 2009; Wiliam, Lee, Harrison, & Black, 2004) with the ECD-CAF framework. The formative assessment framework (Figure 2) is conceived as supporting assessment design by defining the characteristics of assessments that are likely to function in educative / formative ways (Wiliam, 2011).

	Where the learner is going	Where the learner is right now	How to get there
Teachers	1 Clarifying and negotiating learning intentions and criteria for success	2 Engineering effective discussions and other learning tasks, including those using digital media, that elicit evidence of student understanding	3 Providing feedback that moves learners forward
Peers		4 Activating students as instructional resources for one another –peer support, peer feedback	

1/25/13 3:41 PM
 Deleted: s
 1/25/13 3:41 PM
 Deleted: s

Learners

5 Activating students as the owners of their own learning

Figure 2: Aspects of formative assessment adapted from a framework developed by Black and Wiliam (2009)

The Edusummit Assessment Working Group validated the match between the formative assessment framework and their experiences of assessment FOR learning. However, the group also noted that teachers encounter significant challenges in developing both the conditions and methods needed for formative assessment, for example, creating a classroom culture in which students felt able to share their developing understandings (Webb & Jones, 2009). We should not assume therefore that implementing educative / formative assessment in ICT-enabled assessments is a matter only of building on existing classroom practices: instead it will be necessary to work from the conceptual frameworks we have identified and to build on best classroom practice.

The process of clarifying and negotiating learning intentions and criteria involves the students in actively appraising their learning situation and setting targets: a process that needs to be continually revisited. The resulting agreement provides a shared understanding of the student model (in the CAF), the variables of which can be defined as a set of competences, or as tendencies to behave in certain ways in situations with particular features (Mislevy, et al., 2003 P.16).

In both the ECD and formative frameworks evidence is central. In ECD, arguments about evidence are expressed using Toulmin's (1969) scheme of argumentation (Mislevy, et al., 2003). Chains of reasoning about specific assessments are captured in computerised assessments by Bayesian inference networks, to allow for uncertainty, and then fed into the argument (ibid). Statistical methods are used to combine results from a number of assessments. In the formative assessment framework, engagement, understanding and interactive

communications predominate; useful formative feedback requires that evidence-based decisions be made by teachers, peers and learners themselves.

Task design is also important in both frameworks. In the formative assessment framework, task design is focused on eliciting evidence of students' learning with a clear purpose of identifying how to promote their further learning. In the ECD the task model is separate from the student and evidence models thereby allowing for the task to fulfil multiple assessment purposes (Mislevy, Almond, & Lukas, 2004). For example an assessment could provide information to a learner about their performance on one or more tasks as well as contributing to an overall assessment grade based on a range of tasks. The evidence models for these two purposes would be different and the student model would need to be sufficiently fine-grained to provide appropriate detail for the learner.

In summary, frameworks discussed here that come from two very different perspectives: 1) a conceptual approach to assessment design for computerised assessment and 2) empirical research on formative assessment in classrooms, both focus on evidence of student understanding and capabilities. However, while the formative assessment framework has one clear purpose of enabling learning, the ECD can address multiple purposes. Therefore next we consider how purposes of assessments may affect validity.

Reasoning from Evidence and the Purpose of Assessment

A review of the literature (Koch & DeLuca, 2012) suggests that multiple-uses of assessment is neither a recent phenomenon nor a rare occurrence (p.3). However there is general agreement that the purposes of the assessment must be clarified early in the design process in order for an assessment to embody a complete and coherent argument (Mislevy, et al., 2003) and to be valid (Crooks, Kane, & Cohen, 1996). Validity in the ECD process is based on Messick's (1994) construct-centered and Kane's (1992) argument-based approach. An approach to validation

that has been used successfully with teachers (Black, et al., 2010) was based on Crooks et al.'s (Crooks, et al., 1996) chain model in which assessment is divided into eight linked stages (see Table 2). The chain model has the advantages of 1) emphasising that validation is dependent on all eight links so focuses attention on any weak link and 2) potential threats can be discussed by teachers (see examples of threats in Table 2).

TABLE 2. Some threats associated with each of the eight links p. 270 *T. J. Crooks et al*

Link	Threat
Administration	Low motivation Assessment anxiety Inappropriate assessment conditions Task or response not communicated
Scoring	Scoring fails to capture important qualities of task performance Undue emphasis on some criteria, forms or styles of response Lack of intra-rater or inter-rater consistency Scoring too analytic Scoring too holistic
Aggregation	Aggregated tasks too diverse Inappropriate weights given to different aspects of performance
Generalization	Conditions of assessment too variable Inconsistency in scoring criteria for different tasks Too few tasks
Extrapolation	Conditions of assessment too constrained Parts of the target domain not assessed or given little weight
Evaluation	Poor grasp of assessment information and its limitations Inadequately supported construct interpretation Biased interpretation or explanation
Decision	Inappropriate standards Poor pedagogical decisions
Impact	Positive consequences not achieved Serious negative impact occurs

Typically in ongoing classroom assessment for formative purposes teachers only need to consider part of this chain: administration, evaluation, decision and impact. Even in relation to these links the teacher can share the task of validation with the students and thus increase their confidence in their decision. Teachers also use a range of assessments for summative purposes and here they may be making important decisions for reporting and about setting, which should be based on sound reasoning incorporating all links in the validation chain.

However in a detailed study with twelve experienced English and Mathematics teachers, Black et al. (Black, et al., 2010) found that existing practices were weak: teachers did not examine the validity of the assessments that they were using. When teachers were supported to develop assessments their improved understanding of validity also enabled them to engage students and parents in more meaningful discussions of their progress and thus to influence students' motivation (Black, Harrison, Hodgen, Marshall, & Serret, 2011).

These findings and others (e.g. see Koch & DeLuca, 2012; Mansell, et al., 2009; Stanley, MacCann, Gardner, Reynolds, & Wild, 2009) suggest that involving teachers in assessment design is an important element of teaching that can contribute to learning. Furthermore removing this responsibility from teachers through a focus on high-stakes assessment can be damaging to students' learning (Mansell, et al., 2009). Therefore designers of computer-based assessments need to consider how to make the reasoning and decision-making processes transparent so that teachers can be involved in designing, administering and utilizing the results of assessments that provide in-classroom feedback.

The research reviewed here suggests that a focus on validity and reasoning about evidence may support assessment design that may enable the combination of assessment FOR learning and assessment OF learning. However thus far we have only considered using an assessment of an individual student. Using assessments for the wider range of purposes that have recently become commonplace such as evaluating schools and districts and monitoring national performance (Klenowski & Wyatt-Smith, 2011; Mansell, et al., 2009) introduces additional issues for validation (Koch & DeLuca, 2012; Mansell, et al., 2009; Stobart, 2009). For example Koch and DeLuca (Koch & DeLuca, 2012) argue that although existing models of validation provide a sound theoretical base they do not address adequately the complexity of practice, especially for multiple-use of assessment, for two reasons. First, multiple-use may raise the

'stakes' associated with an assessment such that assessments that were formerly low-stakes may become high stakes thus affecting their validity (Kyriakides, 2004; Swift & Spivak, 1969). Second, current validation models assume that assessment uses are independent of one another but multiple-uses are often derived from a single administration of an assessment and their purposes may interact to create unintended consequences (Shaw, Crisp, & Johnson, 2011). Note that validation applies to the assessment in use not to a particular assessment task but if the assessment is designed for multiple uses the potential impact factors and their interactions with other factors would need to be considered simultaneously. For example if students were being assessed through an online game this may be engaging and motivating. If they were encouraged to try hard in order to demonstrate their skills and achieve a badge or some other reward, this may increase their motivation thereby improving the validity in relation to the administration link. However if the school district then decided to use the results to make judgements about the school's performance the assessment would then become high stakes for the school. Therefore the teacher may put pressure on students thus increasing anxiety for some students and changing the validity of the assessment.

Another issue that is particularly important for twenty first century learning (Voogt, 2011) is assessing performance in collaborative learning. The automatic scoring of performance on collaborative tasks is a significant challenge if it is expected to do more than provide a group score. Methods that have been used to make corrections for individual underperformance in group scores (e.g. see Maiden & Perry, 2010) may provide one way forward. In order to achieve validity in scoring of individual performances during a group task, scoring would need to take account not only of individual contributions to completing the task, but also of a range of contextual factors involved in group regulation including motivational, cognitive, metacognitive and socio-emotional regulation and inter-relations between these (e.g. see Järvelä, Volet, & Järvenojä, 2010). These and other factors have been identified by Rupp et al. (2010) specifically in relation to games design where in addition to the validity issues just identified, other factors

concerned with avoiding disturbing the flow of the game need to be accommodated. Otherwise the authenticity and motivational advantages of the game may be lost.

Perhaps the most important validity issue in relation to twenty first century learning is the need for complete construct representation (Messick, 1994) which makes the assessment authentic. In the Crooks et al. chain (Crooks, et al., 1996) incomplete construct representation is a threat in the extrapolation link. Therefore the discussion of this link deserves a broad approach to considering whether the new generation of assessments address the skills and knowledge, in the domain being assessed and in authentic contexts, that are important for twenty first century learning.

How to move towards user-centred assessment enabled by technology

The recognition of the need for more authentic assessment has led on one hand to the development of a new generation of computerized assessment environments (ETS, 2012) and on the other to more teacher involvement in assessment. In a review of teacher assessment worldwide, Stanley et al. (2009) identified an expanded role for teachers in high stakes assessment as well as increased support for teacher assessment: for example in the Australian states of New South Wales and Queensland considerable resources have been devoted to teacher professional development, moderation processes and to creating assessment banks (Stanley, et al., 2009). This broadening of assessment is based on a view that there are aspects of learning that are important but cannot be adequately assessed by formal external tests. These aspects require human judgment to integrate the many elements of performance behaviours that are required in dealing with authentic assessment tasks (Stanley et al. 2009, p.31)

In order to address the need to involve users Nichols et al. (2008) proposed combining elements of User Centred Assessment Design (UCAD) into ECD. User Centred Design (UCD) on which UCAD is based has an obvious appeal: no-one would want to suggest that users were

not central to a design process. In practice however UCD is variously interpreted, so a more practical approach may be to reflect upon how any systems development method may be made user-centred (livari & livari, 2011). Thus, as discussed earlier in relation to ECD, a more user-centred approach would ensure that all parties (teachers, students, others) develop a shared understanding of reasoning from evidence and discussing validity within the contexts of decisions and judgments based on the assessment. This might enable, for example, teachers and learners to select tasks from a database and combine them into an assessment to suit their particular needs. Learners could opt to choose from a range of tasks to demonstrate their capabilities. An assessment system to be used in this way must be able to adaptively present to users, in meaningful ways, how its reasoning about performance on these tasks would demonstrate their capabilities. Furthermore the system should enable users to input assessment results with tagging for validity so that aspects of achievement that still require assessments by humans can be incorporated into assessment evidence.

In order to support learners, assessment systems need to incorporate effective feedback processes. However as Perrenoud (1998) argued, feedback 'can only be effective if a window is found into the cognitive system of the learner' (p. 88). Depending on the particular learner, their needs and a range of contextual factors, this 'window' might be found by providing a simple statement as rapid corrective feedback (Hattie & Temperley, 2007) or it may require extensive teacher or peer interaction as discussed earlier. For a computer system to be capable of the range of interactions that experienced teachers adaptively engage in with their pupils is some way in the future. However there are immediate benefits to be gained from computerised assessment systems analysing the frequency and nature of student misunderstandings and feeding them into professional development discussions (Nichols, et al., 2008). Furthermore, while fully automated feedback systems are challenging to develop there are also many

opportunities to use relatively simple ICT tools, including Web 2.0 applications, to support peer assessment and feedback (Webb 2011).

Thus combining the ECD and formative assessment frameworks and building on the opportunities provided by computerised assessments, as well as harnessing teachers' experience and developing their validation processes through discussion, is proposed for enabling assessments to address all four of the possibilities depicted in Table 1. In this way the same assessment could embrace assessment FOR learning and assessment OF learning provided that designers and users of assessment are mindful of the dangers of over-emphasis on Perspective 4 about making summative judgements. Specifically we recommend that summative judgements on individuals are made only at important transition points where reviewing each learner's individual performance is essential for reporting, accreditation or setting. Summative judgements for other more general evaluative purposes such as comparing cohorts, schools, etc. should be researched carefully and tightly controlled in order to avoid uninformed administration regimes adverse effects on validity and students' futures.

CONCLUSIONS AND IMPLICATIONS

The analysis of recent research presented in this paper has confirmed the potential for IT to play a major role in transforming assessment practices to support the needs of learners as well as the needs of educational systems (e.g. classroom practices, school organizations, national priorities), in the twenty first century. A number of challenges have also been identified and some ways of addressing them have been explored.

[Evidence-Centered Design](#), provides a useful framework for creating components that could be re-assembled fairly easily to provide assessments that address users' needs (Mislevy, et al., 2003). However in order for users to benefit fully, developers need to consider how to make the reasoning and decision-making processes accessible to users so that they not only understand them but can contribute to assessment development and interpretation. There is thus a need to

1/25/13 3:46 PM

Deleted: CD

develop assessment literacy (Stiggins, 1995) in teachers and other users so that they understand the advantages and limitations of assessment types and processes and are confident in developing and analysing arguments from evidence based on current understanding of validation (Black, et al., 2010). These processes of professional development could also be supported by computerised assessment systems that analyse data about student performance and misunderstandings. In this way development could become a shared process between assessment bodies, teachers and other stakeholders.

Developments of digitally-enhanced assessments are at an early stage but they have the potential to increase the range of assessment measures thus contributing to complete construct representation and increasing authenticity of assessment. Improvements in computer supported statistical analysis promise to make the analysis of the enormous quantities of data generated, more manageable. However as we have argued in this paper, significant challenges remain for developing validation approaches that can take account of the complexity of these learning experiences especially for group tasks in simulations, games and other problem solving environments. If these challenges can be overcome so that digitally-enhanced assessments through simulations and games can become important components of high stakes assessments, they offer the promise of overcoming some of the negative impacts of current high stakes assessments, for example, on the development of creativity and social development (Harlen and Deakin Crick 2002).

The benefits of embedded continuous unobtrusive measuring of performance are obvious in that learners can become engaged in interesting tasks that have been designed to support their learning. However, at the same time, as we have argued in this paper, learners also need to be involved in discussing and negotiating their learning intentions so they should have access to meaningful representations of arguments about their achievements. Therefore rather than describing this unobtrusive measuring as “stealth assessment,” which has connotations of

secrecy and furtiveness, we prefer to conceptualise these processes as “quiet assessment” whose volume can be turned up by learners and teachers whenever they wish.

As outlined above, our analysis of existing research has enabled us to identify ways in which the same assessment could engage the students in meaningful activity, contribute to feedback information, improve decisions to support learning and enable judgements about knowledge and skills that are shared with learners and teachers. Such approaches would combine formative and summative assessment purposes to the benefit of learners while at the same time helping educational environments to continuously improve by, for example, setting and monitoring targets for individual students and groups and using aggregated data on student misunderstandings to inform curriculum planning. However even with the increased possibilities that ICT provides we have not yet found a way to say confidently that the multiple purposes for which some assessments have been used (Mansell, et al., 2009) can or should be supported through the same assessment systems. This is because the impacts of some purposes interact with the validation processes for others. Therefore in considering assessment design for multiple purposes users need to examine impact factors carefully in order to minimise negative impacts on learning and learners.

This paper has identified ways of moving assessment design towards a more user-centred approach and particularly of meeting learners’ needs. The paper has focused predominantly on school-based assessments because school is where most young learners are assessed even though much of their learning happens outside of school in various informal settings. Portfolios have been recommended by various researchers to encompass the need for assessing a greater range of performance (Black, et al., 2010; Moss, 1994) and they have been used successfully to support high-stakes assessments where criteria have been identified clearly (Stanley, et al., 2009). The assessment approaches discussed in this paper could be applied to the design of portfolios (see for example Gibson, 2010) that could incorporate not only evidence

from school-based assessments but also from other environments and activities in which learners are involved. In these ways learners needs, as they change and diversify in the twenty first century, may be met.

ACKNOWLEDGMENTS

(acknowledge the contributions of the members of TWG 5)

REFERENCES

- BENNETT, R. E. (2010). COGNITIVELY BASED ASSESSMENT OF, FOR, AND AS LEARNING (CBAL): A PRELIMINARY THEORY OF ACTION FOR SUMMATIVE AND FORMATIVE ASSESSMENT. *MEASUREMENT: INTERDISCIPLINARY RESEARCH AND PERSPECTIVES*, 8(2-3), 70-91.
- BLACK, P., HARRISON, C., HODGEN, J., MARSHALL, B., & SERRET, N. (2010). VALIDITY IN TEACHERS' SUMMATIVE ASSESSMENTS. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 17(2), 215-232.
- BLACK, P., HARRISON, C., HODGEN, J., MARSHALL, B., & SERRET, N. (2011). CAN TEACHERS' SUMMATIVE ASSESSMENTS PRODUCE DEPENDABLE RESULTS AND ALSO ENHANCE CLASSROOM LEARNING? *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 18(4), 451-469.
- BLACK, P., HARRISON, C., LEE, C., MARSHALL, B., & WILIAM, D. (2002). *WORKING INSIDE THE BLACK BOX: ASSESSMENT FOR LEARNING IN THE CLASSROOM*. LONDON: KING'S COLLEGE, LONDON, DEPARTMENT OF EDUCATION & PROFESSIONAL STUDIES.
- BLACK, P., HARRISON, C., LEE, C., MARSHALL, B., & WILIAM, D. (2003). *ASSESSMENT FOR LEARNING: PUTTING IT INTO PRACTICE*. BUCKINGHAM, UK: OPEN UNIVERSITY.
- BLACK, P., & WILIAM, D. (2009). DEVELOPING THE THEORY OF FORMATIVE ASSESSMENT. *EDUCATIONAL ASSESSMENT, EVALUATION AND ACCOUNTABILITY*, 21, 5-31.
- BUTLER, R. (1988). ENHANCING AND UNDERMINING INTRINSIC MOTIVATION; THE EFFECTS OF TASK-INVOLVING AND EGO-INVOLVING EVALUATION ON INTEREST AND PERFORMANCE. *BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY*, 58, 1-14.
- CLARKE, J., & DEDE, C. (2010). ASSESSMENT, TECHNOLOGY, AND CHANGE. *JOURNAL OF RESEARCH IN TEACHER EDUCATION*, 42(3).
- CROOKS, T. J., KANE, M. T., & COHEN, A. S. (1996). THREATS TO THE VALID USE OF ASSESSMENTS. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 3(3), 265-285.
- ERSTAD, O. (2008). CHANGING ASSESSMENT PRACTICES AND THE ROLE OF ICT. IN J. KNEZEK & J. VOOGT (EDS.), *INTERNATIONAL HANDBOOK OF INFORMATION TECHNOLOGY IN EDUCATION* (PP. 163-180). NEW YORK.
- ERTMER, P. A., RICHARDSON, J. C., BELLAND, B., CAMIN, D., CONNOLLY, P., COULTHARD, G., ET AL. (2007). USING PEER FEEDBACK TO ENHANCE THE QUALITY OF STUDENT ONLINE POSTINGS. *JOURNAL OF COMPUTER-MEDIATED COMMUNICATION*, 12 (2), 1-15.
- FORKOSH-BARUCH, A., GIBSON, D., SCHULZ-ZANDER, R., & WEBB, M. (2009). *ICT IN TEACHING AND LEARNING*. THE HAGUE, NL: EDUSUMMIT 2009.
- GIBSON, D. (2010). ASSESSMENT AND DIGITAL MEDIA LEARNING. RETRIEVED FROM [HTTP://PREZI.COM/14BPZ9TG3T2J/](http://PREZI.COM/14BPZ9TG3T2J/)
- GIBSON, D. (2011). *MODELING EMOTIONS IN SIMULATED LEARNING*. *STANDARDS IN EMOTION MODELING*: LORENTZ CENTER INTERNATIONAL CENTER FOR WORKSHOPS IN THE SCIENCES.
- GIBSON, D., ALDRICH, C., & PRENSKY, M. (2007). *GAMES AND SIMULATIONS IN ONLINE LEARNING: RESEARCH AND DEVELOPMENT FRAMEWORKS*: INFORMATION SCIENCE PUBLISHING.

- HARLEN, W., & JAMES, M. (1997). ASSESSMENT AND LEARNING: DIFFERENCES AND RELATIONSHIPS BETWEEN FORMATIVE AND SUMMATIVE ASSESSMENT. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 4(3), 365-379.
- HATTIE, J., & TEMPERLEY, H. (2007). THE POWER OF FEEDBACK. *REVIEW OF EDUCATIONAL RESEARCH*, 77(1), 81-112.
- IIVARI, J., & IIVARI, N. (2011). VARIETIES OF USER-CENTREDNESS: AN ANALYSIS OF FOUR SYSTEMS DEVELOPMENT METHODS. *INFORMATION SYSTEMS JOURNAL*, 21(2), 125-153.
- JÄRVELÄ, S., VOLET, S., & JÄRVENOJÄ, H. (2010). RESEARCH ON MOTIVATION IN COLLABORATIVE LEARNING: MOVING BEYOND THE COGNITIVE-SITUATIVE DIVIDE AND COMBINING INDIVIDUAL AND SOCIAL PROCESSES. *EDUCATIONAL PSYCHOLOGIST*, 45(1), 15-27.
- KANE, M. T. (1992). AN ARGUMENT-BASED APPROACH TO VALIDITY. *PSYCHOLOGICAL BULLETIN*, 112, 527-535.
- KLENOWSKI, V., & WYATT-SMITH, C. (2011). THE IMPACT OF HIGH STAKES TESTING: THE AUSTRALIAN STORY. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 1-15.
- KOCH, M. J., & DELUCA, C. (2012). RETHINKING VALIDATION IN COMPLEX HIGH-STAKES ASSESSMENT CONTEXTS. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 1-18.
- LENHARD, W., BAIER, H., HOFFMANN, J., & SCHNEIDER, W. (2007). AUTOMATIC SCORING OF CONSTRUCTED-RESPONSE ITEMS WITH LATENT SEMANTIC ANALYSIS. *DIAGNOSTICA*, 53(3), 155-165.
- MAIDEN, B., & PERRY, B. (2010). DEALING WITH FREE-RIDERS IN ASSESSED GROUP WORK: RESULTS FROM A STUDY AT A UK UNIVERSITY. *ASSESSMENT & EVALUATION IN HIGHER EDUCATION*, 36(4), 451-464.
- MANSELL, W., JAMES, M., & GROUP, T. A. R. (2009). *ASSESSMENT IN SCHOOLS. FIT FOR PURPOSE? A COMMENTARY BY THE TEACHING AND LEARNING RESEARCH PROGRAMME*. LONDON: ECONOMIC AND SOCIAL RESEARCH COUNCIL:TEACHING AND LEARNING RESEARCH PROGRAMME.
- MESSICK, S. (1994). THE INTERPLAY OF EVIDENCE AND CONSEQUENCES IN THE VALIDATION OF PERFORMANCE ASSESSMENTS. *EDUCATIONAL RESEARCHER*, 23(2), 13-23.
- MISLEVY, R. J., ALMOND, R. G., & LUKAS, J. F. (2004). *A BRIEF INTRODUCTION TO EVIDENCE-CENTERED DESIGN (CSE REPORT 632)*: CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING.
- MISLEVY, R. J., STEINBERG, L. S., & ALMOND, R. G. (2003). ON THE STRUCTURE OF EDUCATIONAL ASSESSMENT. *MEASUREMENT: INTERDISCIPLINARY RESEARCH AND PERSPECTIVE*, 1(1), 3-62.
- MOSS, P. A. (1994). CAN THERE BE VALIDITY WITHOUT RELIABILITY? *EDUCATIONAL RESEARCHER*, 23(2), 5-12.
- NICHOLS, P., MITTELHOLTZ, D., ADAMS, J., & VAN DEUSEN, R. (2008). *USER-CENTERED ASSESSMENT DESIGN*. PAPER PRESENTED AT THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION.
- PERRENOUD, P. (1998). FROM FORMATIVE ASSESSMENT TO A CONTROLLED REGULATION OF LEARNING PROCESSES. TOWARDS A WIDER CONCEPTUAL FIELD *ASSESSMENT IN EDUCATION*, 5(1), 85-102.
- RESTA, P., SEARSON, M., PATRU, M., KNEZEK, G., & VOOGT, J. (2012). *BUILDING A GLOBAL COMMUNITY OF POLICY-MAKERS, RESEARCHERS AND TEACHERS TO MOVE EDUCATION SYSTEMS INTO THE DIGITAL AGE: SUMMARY REPORT OF EDUSUMMIT 2011 AND CALL TO ACTION*: EDUSUMMIT 2011.

- RISSANEN, M. J., KUME, N., KURODA, Y., KURODA, T., YOSHIMURA, K., & YOSHIHARA, H. (2008). ASYNCHRONOUS TEACHING OF PSYCHOMOTOR SKILLS THROUGH VR ANNOTATIONS: EVALUATION IN DIGITAL RECTAL EXAMINATION. *STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS*, 132, 411-416.
- RUPP, A. A., GUSHTA, M., MISLEVY, R. J., & SHAFFER, D. W. (2010). EVIDENCE-CENTERED DESIGN OF EPISTEMIC GAMES: MEASUREMENT PRINCIPLES FOR COMPLEX LEARNING ENVIRONMENTS. *THE JOURNAL OF TECHNOLOGY, LEARNING, AND ASSESSMENT VOLUME*, 8(4).
- SABOURIN, J., MOTT, B., & LESTER, J. (2011). *COMPUTATIONAL MODELS OF AFFECT AND EMPATHY FOR PEDAGOGICAL VIRTUAL AGENTS*. LEIDEN, NL: LORENTZ CENTER INTERNATIONAL CENTER FOR WORKSHOPS IN THE SCIENCES.
- SHUTE, V. J. (2011). STEALTH ASSESSMENT IN COMPUTER-BASED GAMES TO SUPPORT LEARNING. IN S. TOBIAS & J. D. FLETCHER (EDS.), *COMPUTER GAMES AND INSTRUCTION* (PP. 503-524). CHARLOTTE, NC: INFORMATION AGE PUBLISHERS.
- STANLEY, G., MACCANN, R., GARDNER, J., REYNOLDS, L., & WILD, I. (2009). *REVIEW OF TEACHER ASSESSMENT: EVIDENCE OF WHAT WORKS BEST AND ISSUES FOR DEVELOPMENT*. OXFORD CENTRE FOR EDUCATIONAL ASSESSMENT, OXFORD UNIVERSITY.
- STIGGINS, R. J. (1995). ASSESSMENT LITERACY FOR THE 21ST CENTURY. *PHI DELTA KAPPAN*, 77.
- STIGGINS, R. J. (2005). *STUDENT-INVOLVED ASSESSMENT FOR LEARNING (4TH ED.)*. UPPER SADDLE RIVER, N.J: MERRILL/PRENTICE HALL.
- STOBART, G. (2009). DETERMINING VALIDITY IN NATIONAL CURRICULUM ASSESSMENTS. *EDUCATIONAL RESEARCH*, 51(2), 161-179.
- TOULMIN, S. (1969). *THE USES OF ARGUMENT*. CAMBRIDGE, ENGLAND: CAMBRIDGE UNIVERSITY PRESS.
- TURKAY, S., & TIRTHALI, D. (2010). YOUTH LEADERSHIP DEVELOPMENT IN VIRTUAL WORLDS: A CASE STUDY. *PROCEDIA SOCIAL AND BEHAVIORAL SCIENCES*, 2(2), 3175-3179.
- VAN DER POL, J., VAN DEN BERG, B. A. M., ADMIRAAL, W. F., & SIMONS, P. R. J. (2008). THE NATURE, RECEPTION, AND USE OF ONLINE PEER FEEDBACK IN HIGHER EDUCATION. *COMPUTERS & EDUCATION*, 51(4), 1804-1817.
- VOOGT, J. (2011). *TEACHER COMPETENCIES FOR 21ST CENTURY PEDAGOGY*. PAPER PRESENTED AT THE SOCIETY FOR INFORMATION TECHNOLOGY & TEACHER EDUCATION INTERNATIONAL CONFERENCE 2011.
- VOOGT, J., & KNEZEK, G. (2008). *INTERNATIONAL HANDBOOK OF INFORMATION TECHNOLOGY IN PRIMARY AND SECONDARY EDUCATION*. LONDON: SPRINGER.
- WEBB, M. E. (2010). BEGINNING TEACHER EDUCATION AND COLLABORATIVE FORMATIVE E-ASSESSMENT. *ASSESSMENT AND EVALUATION IN HIGHER EDUCATION*, 35(5), 597 - 618.
- WEBB, M. E., & GIBSON, D. (2011). *ASSESSMENT TO MOVE EDUCATION INTO THE DIGITAL AGE: BRIEF REPORT FROM THEMATIC WORKING GROUP TWG 5 ON ASSESSMENT*. PAPER PRESENTED AT THE EDUSUMMIT 2011: BUILDING A GLOBAL COMMUNITY OF POLICY-MAKERS, EDUCATORS AND RESEARCHERS TO MOVE EDUCATION INTO THE DIGITAL AGE. RETRIEVED FROM [HTTP://EDUSUMMIT.NL/RES2011/CALLTOACTION2011/BRIEFPAPERS2011](http://edusummit.nl/res2011/calltoaction2011/briefpapers2011)
- WEBB, M. E., & JONES, J. (2009). EXPLORING TENSIONS IN DEVELOPING ASSESSMENT FOR LEARNING. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 16 (2), 165-184.

- WIGGINS, G. (1999). *EDUCATIVE ASSESSMENT: DESIGNING ASSESSMENTS TO INFORM AND IMPROVE STUDENT PERFORMANCE*. SAN FRANCISCO: JOSSEY BASS PUBLISHERS.
- WILIAM, D. (2011). WHAT IS ASSESSMENT FOR LEARNING? *STUDIES IN EDUCATIONAL EVALUATION*, 37(1), 3-14.
- WILIAM, D., LEE, C., HARRISON, C., & BLACK, P. (2004). TEACHERS DEVELOPING ASSESSMENT FOR LEARNING: IMPACT ON STUDENT ACHIEVEMENT. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY AND PRACTICE*, 11(1), 49-65.
- YEH, S. S. (2010). UNDERSTANDING AND ADDRESSING THE ACHIEVEMENT GAP THROUGH INDIVIDUALIZED INSTRUCTION AND FORMATIVE ASSESSMENT. *ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE*, 17(2), 169-182.