Report on current state of the art in formative and summative assessment in IBE in STM - Part II

Sarah Hughes
Clare Green
Vanessa Greene

ASSISTME   **Assess Inquiry in Science, Technology and Mathematics Education**

Assess Inquiry in Science, Technology and Mathematics Education

ASSISTME

Report on current state of the art in formative and summative assessment in IBE in STM – Part II: Digital assessments

Sarah Hughes, Clare Green and Vanessa Greene

(for Pearson Education International)

Table of Contents

1 Background and context

1.1   The ASSIST-ME project

The *Assess Inquiry in Science, Technology and Mathematics Education (ASSIST-ME)* project is an EU funded Europe-wide project which aims to investigate formative and summative assessment methods to support and to improve inquiry-based approaches in European science, technology and mathematics (STM) education.

A number of work packages will constitute the project which began in early 2013 and will be complete in 2017.

Work package	Title
1	Project Management
2	Synthesise existing research
3	Characterise Educational Systems
4	Design Assessment Methods
5	Trial Implementation of Assessment Methods
6	Transform Results into National Contexts
7	Promotion of Guidelines and Results

Based on an initial analysis of the literature (WP2), to identify what is known about summative and formative assessment related to key STM competences and IBE, and an analysis of European educational systems (WP3), the project will design a range of assessment methods (WP4). These methods will be tested as part of the project in primary and secondary schools in different educational cultures in Europe (WP5). This will enable an analysis of the conditions that support or undermine the uptake of formative and summative assessment related to inquiry processes (WP6).

Reflections on the development and trialling of the assessments will enable the formulation of guidelines and recommendations for policy makers, curriculum developers, teacher trainers and other stakeholders in the different European educational systems (WP7).

1.2   Work Package 2 – the literature review

A number of work packages (WPs) make up the project. WP2 comprises a literature review which aims to analyse existing research on how summative and formative assessment of knowledge, competences and attitudes in STM can be coupled with inquiry-based teaching. Pearson Education's role in that literature review is to review the use of e-assessment in the formative and summative assessment of STM subjects at primary and secondary levels with a focus on inquiry-based and competence-based learning.

Pearson Education's work on WP2 is an independent forerunner for Pearson's work on WPs 4 and 5 (the development and trialling of assessments). This independence will ensure that the best practice identified in the literature review is reported and implications for the development of e-assessments are described, without reference to what might be possible or desirable, from Pearson's point of view, to propose as a design for the ASSIST-ME e-assessments.

1.3   The purpose and objectives relating to this report

This report is the output of Pearson's work in WP2.

The purpose of this report is to provide a literature review that will inform the development of digital assessments which are relevant to the aims of ASSIST-ME, that is, it will:

1. enable both formative and summative assessment

2. cover STM subjects

3. focus on inquiry-based education

4. focus on competence-based learning

5. be relevant to primary and/or secondary education.

To enable us to fulfil that purpose, our objectives were to

1. Through the literature, identify theories and models which are relevant to the development of such digital assessments

2. Identify strategies used in the evaluation of the models which could inform good practice

3. Identify existing relevant digital assessments

4. Identify implications for the development of the digital assessments relevant to the aims of ASSIST-ME.

This report contains sections on the methodology employed followed by the findings, described under themed heading including implications for work packages 4 and 5: the devel-

opment of e-assessments for ASSIST-ME. The conclusions are structured around the four objectives listed above.

2 Methodology

2.1 E-assessment

The concepts of competence, inquiry-based STM education and assessment are all described in the ASSIST-ME project proposal, where definitions of terms are shared. We would like to clarify what we mean by e-assessment in the context of the ASSIST-ME project because this impacts on the scope and implications of the work. Beevers and Winkley (2011) provide two definitions of e-assessment, one focuses on the e-administration of tests and the other one relates to providing automation in the pedagogic process:

  A. e-assessment occurs when there is an automated marking/response to student input on-screen in a test, informing on the process of answering a question and providing feedback to learners and their teachers through well-crafted advice and reports.

Alternatively,

  B. e-assessment occurs when there is use of technology in testing which encompasses the on-screen computer-marked assessments of (A) above but also includes on-screen human marking of tests, electronic management and presentation of results, moderation and awarding processes with awarding bodies, anti-plagiarism software, tools which enable collaboration on the assessment and feedback processes, voting systems/clickers and e-portfolios.

This difference in scope is not helpful as it confuses those who are not 'into e-assessment' and even allows experts to talk at crossed purposes at times.

JISC (2007) gives an alternative definition of e-assessment: The range of activities in which digital technologies are used in assessment – designing and delivering assessments, marking, processes of reporting, storing and transferring data. More recently, Broadfoot et al. (2013a) explored technology-enhanced assessment which refers to the wide range of ways in which technology can be used to support assessment and feedback and includes on-screen or e-assessment.

For the purpose of this literature review e-assessment was taken to include:

- the onscreen presentation of tasks and tests

- delivery of assessments

- automated marking

- automated feedback to students

- students' onscreen and digital responses

- creation, management and manipulation of data for teachers and

- tools which enable collaboration on the assessment and feedback processes.

2.2   Sources

Initial searches led to a list of productive journals. These, and others, were searched using the terms described in section 2.3 and Appendix 1, covering at least the last 10 years. Searches were not confined geographically; all countries were included, but only those sources which were available in the English language were considered.

Priority journals:

1. Educational Technology, Research and Assessment

2. Journal of Technology, Learning, and Assessment

3. British Journal of Educational Technology

Other journals:

4. Computer-Based Testing

5. Computers and Education

6. Education and Information Technologies

7. European Journal of Education: special issue – ICT and Education

8. Frontiers in Artificial Intelligence and Information and Communication Technologies

9. International Encyclopaedia of Education (Technology and Learning - assessment)

10. International Journal of Computer-Supported Collaborative Learning

11. International Journal of E-assessment (Journal of the E-assessment Association)

12. International Journal of Educational Research

13. Journal of Applied Testing Technology

14. Journal of Computer Assisted Learning

15. Journal of Information Technology in Teacher Education

16. Journal of Research on Computing in Education

17. Journal of Science Education and Technology

18. Learning, Media and Technology

19. Research in Learning Technology (Journal of the Association of Learning Technology)

Other types of sources that were searched included:

- Organisations (e.g. NFER, BECTA, OECD)

- Government websites (e.g. Ofqual, Office for Official Publications of the European Communities)

- Specific assessment projects and examples of online assessments (e.g. Operation ARIES!)

- Conferences (e.g. Computer Aided Learning Conference, International Conference on Intelligent Tutoring Systems)

- University departments (e.g. The Centre for Mathematics, Science and Computer Education, Rutgers University)

2.3   Search terms

All searches targeted e-assessment (which incorporated e-learning). We searched for instances where e-assessment occurred as a key word alongside these other keywords or their alternatives shown in Appendix 1:

– formative assessment

– summative assessment

– inquiry based education

– competences

– mathematics

– science

– technology.

3 Findings

Having applied the methods, approximately eighty articles and sources were included in this literature review. Emerging from the literature was a number of interrelated themes relevant to the use of e-assessment in the ASSIST-ME project. A visual representation of these is presented below, followed by a description of findings for each theme and sub-theme. This visual representation does not do justice to the extent of overlap between the themes. Given this overlap between themes, it is possible that the findings could have been organised in a number of different ways, but these themes and their order were chosen to best fit with the aims and characteristics of the ASSIST-ME assessments.

The themes under the heading 'Aspects of e-assessment tasks/items' provide an overview of e-assessment and its component parts. This is given as a background and framework on which the reader can hang the detail contained in the rest of the findings.

Assessment / teaching and Learning link (Section 3.2)

Examples of intervention

Resubmitting issues

Intelligent tutoring
Animated objects
Adaptive assessment
Send to learning object
Demonstration answers

**Inquiry based education
(Section 3.3)**

**Competence based learning
(Section 3.4)**

Formative assessment (Section 3.5)

Self-assessment

Peer assessment

Diagnostic assessment

Teacher / learner / computer relationship

Confidence ratings

Group assessment

Misconceptions

Prompts for discussion

**Summative assessment
(Section 3.6)**

**Formative / summative assessment
link (Section 3.7)**

Feedback (Section 3.8)

Output to teacher	Feedback to learner
Learning analytics *Feedback level*	*Characteristics of feedback: type, timing, level of detail, frequency*

Effects of e-assessment on the learner (Section 3.9)

Motivation	Learning gains	Confidence

Quality (Section 3.10)

Value added paper to e-assessment	Validity	Reliability

Implications for the implementation of e-assessment (section 3.11)

Resources and technical equipment	CPD teacher training	Impetus for e-assessment	Time / frequency of formative assessment
	Use of technology *Use of e-assessment* *Formative assessment* *FA / SA link* *IBE*		

ASSIST

Implications for evaluation of assessments (Section 3.12)

Exemplars (Section 3.13)

| Science | Mathematics | Technology |

3.1 Aspects of e-assessment tasks/items

An e-assessment (like a paper-based assessment) comprises a stimulus, the student response, some form of feedback and, potentially, an intervention. A description of these aspects, and examples taken from e-assessments identified in the literature follow.

3.1.1 E-Assessment type

3.1.1.1 Replication of paper test

Hughes et al. (2011) in a comparison of paper and onscreen mathematics assessment questions found no questions that were more accessible onscreen than on paper. It was predicted that questions in which the response mode was ordering or involved dragging and dropping objects would favour computer-delivery and performance would be higher onscreen, but this was not the case. Two particular question characteristics which were more accessible on paper were questions which required annotation and questions of a visual nature. It was also the case, although less strongly, that graphical questions and dense questions performed differently on paper and onscreen, with a bias towards paper. This is likely to be because annotating and working with diagrams is more difficult when working on screen, as there were no resources to allow students to interact with the visuals and diagrams as those working on paper could.

3.1.1.2 Gaming style assessments

Gee and Schaffer (2010) promoted the use of gaming for problem solving: 'Video games are good for learning because games can create virtual worlds where players solve simulations of real-world problems and in the process learn real-world skills, knowledge and values.' They added that the choices pupils make while problem solving can tell teachers a great deal about their ability to learn new material later on. Games involve continual diagnostic assessment of strengths and weaknesses in thinking, giving a portrait of problem solving decisions over time, so feedback to customise learning can be provided. They did highlight an interesting issue about the priorities when designing games compared to assessments. They observed that games are constructed in the opposite way to education: games consider first how to test and challenge a player and then design the learning, whereas in educational contexts we are more used to designing the learning and then the assessment to reflect the learning goals. Kennewell (2008) reported that the success of an e-assessment system may be influenced by the extent to which the software pre-specifies course activity. This suggests that a shift to the gaming approach, whereby the tools available and their affordances lead e-assessment content and processes, can lead to better e-assessments.

Games focus on problem solving with a mix of practice and guidance, complex concepts are introduced when needed and when a player's position in the game suggests that they would be most beneficial. Typically in gaming, players spend a lot of time on task, and are motivated because they are presented with a sequence of activities gradually increasing in difficulty, which means that players are constantly working at the edge of their abilities (Gee and Schaffer, 2010).

Johnson (2013) described an emerging field where computer game development and educational assessment are coming together, in which games are able to capture valid and reliable

evidence. Both seek to engage learners by placing them in situations where they face challenges. Games then provide feedback in response to the choices made while pushing capabilities to the limit, rendering games similar to adaptive tests moving along different paths depending on skills. Good e-assessment extracts data not only from the results of test takers, but also from the processes they used to achieve those results (e.g. the use of gaming technology can provide evidence of what actions the learner takes during an activity, what they learn). As is good practice for developing any assessment, in whatever mode, developers should focus on the kinds of knowledge and skills they want the learners to display in an assessment.

Operation ARA is an example of a gaming style assessment (Halperna et al. 2012) and is described in section 3.13.2.2.

3.1.2   Stimulus type

3.1.2.1   Real life scenario

Koenig (2011) described Operation ARIES!, a scenario based e-learning tool for science. Real life examples are used to help students transfer what they have learnt in one context to another scenario-based assessment, and this was argued to be useful where there was a need to apply knowledge to practical situations. Koenig also described aspects of Operation ARIES! where students were expected to apply what they have learnt in the previous modules, for example, students were presented with inaccurate science information through the medium of newspaper headlines and television news channels and had to ask questions to ascertain the truth.

Halperna et al. (2012) furthered the work on Operation ARIES! (renamed Operation ARA), retaining the valued scenario-based assessments, in which students applied their understanding of scientific concepts to determine whether a described research case was reliable or flawed.

3.1.2.2   Simulation

Simulations enable students to interact with and control variables using technology. Simulations can function as independent learning tools, but are also valuable for assessment purposes (Neumann, 2010). Simulations can superimpose multiple representations and permit manipulation of structures and patterns that otherwise might not be visible, they can probe knowledge of how components of a system interact, as well as encourage learners to investigate the impact of varying multiple variables simultaneously (Quellmalz, 2009).

Examples of simulation use were found in all three STM subjects:

Science Clesham (BERA, 2009) developed and trialled secondary science, computer-based simulations to teach and assess scientific enquiry skills. Interactive investigations were developed by storyboarding investigative processes. It included experimentation using interactive simulators (modelling trialling and data collection) and questions involving manipulation of on-screen tools.

Pellegrino and Quellmalz's (2010) paper on 'SimScientists' an online environment for teaching and learning in science, illustrates ways that assessment tasks can take advantage of the benefits of simulations to represent generalisable, progressively complex models of sci-

ence systems and such innovative items were included in the 2009 NAEP science administration.

Mathematics Neumann (2010) used a statistics simulation (followed by multiple choice questions) for summative assessment in HE.

Technology The new 2014 Technology and Engineering Literacy Framework for NAEP will be entirely computer administered and will include specifications for interactive, simulation-based tasks involving problem solving, communication, and collaboration related to technology and society, design and systems, and information communications technology (Pellegrino and Quellmalz, 2010). In the UK, a national test in ICT for 14-year olds was piloted, which involved the simulation of a desktop with a suite of programmes and email software (Boyle, 2006).

3.1.2.3 Interactive

Interactivity often occurs alongside the use of simulations in technology enhanced learning and assessment.

Neumann (2010) developed an interactive statistics simulation used for summative assessment. Neumann described how the affordances of the technology were exploited in the design of the tasks: 'The crucial aspect of each simulation was that it was interactive. Students were able to change data values, simulate events, and see what effects their changes had. It was this interactive nature that was exploited in the assessment approach.' (Neumann, 2010)

Beevers et al. (2011) also valued the use of interactivity; through the CALM project lessons were learned including how to design assessments to give learners more autonomy through interactivity. Another example is Operation ARIES! which was designed to assess and teach critical thinking about science (Koenig, 2011). It uses intelligent tutoring and makes use of animated characters. Students receive feedback and tutorage from two characters from the program throughout. Here the interactive element is how the learner is able to interact with virtual peers and tutors.

3.1.2.4 Video

Many e-assessments use multimedia stimuli followed by objective questions, for example, Operation ARIES! which aimed to engage students by using multimedia; it was designed to assess and teach critical thinking about science. It uses intelligent tutoring and makes use of animated characters. Students watch videos and receive communications through email and text message (Koenig, 2011).

3.1.2.5 Examples

Animation has also been used to improve examples given at the beginning of a test or set of items to show students how to answer questions. Direct observation in class suggests that many pupils rush straight into answering the questions in a worksheet without reading examples. Animating the examples makes them more interesting and encourages learners to engage with them. In some e-assessments, engagement of the learner with the example answers is encouraged by the example remaining on screen for a set time, with the learner not being able to move on from that screen until this time is up. Animated examples may also

hold the attention of the learner more than static ones (onscreen or on paper). Some systems now also ask students how well they have understood the example before proceeding on to the questions. In principle, their answers could inform future navigation.

3.1.3 Response type

3.1.3.1 MCQ and short answer question

There were concerns raised within the literature as to whether multiple choice questions (MCQs) are suitable for computer based formative assessment (e.g. Velan et al., 2008). Although Velan et al. found that MCQs used in formative assessments did have a positive impact of the learning of medical undergraduates, they recognised that this was a different outcome to most other research.

The literature described two key ways in which MCQs have been used productively in formative assessments:

1. MCQs or other objective questions have been presented to students after they have worked on a simulation or interactive stimulus.

2. Crisp and Ward (2008) used objective item types following a simulation of a real life situation (in the context of teacher training). Feedback gave the correct answer(s) with reasons and, where necessary, explained why the student's choice was not correct. E.g. within Operation ARIES! Koenig (2011) and the subsequent tool Operation ARA (Halperna et al., 2012) reported that during a phase of e-learning, learners read an e-book and after each chapter they were quizzed with multiple choice-type questions

MCQs can also provide useful outputs when responses, both correct and erroneous, are codified to allow the individualisation of feedback. After the MCQs are attempted in Operation ARIES! students participated in trialog discussions with avatars where the understanding of the material from the chapter is clarified and reinforced.

Wylie and Dolan (2013) raised the issue that when using MCQs in technology enhanced formative assessment, the challenge is in supporting teachers to use outcomes to stimulate discussion and move understanding forward. Wylie and Dolan used MCQs specifically to identify any mathematical misconceptions held by secondary school students and output these to the teacher. Wylie suggested particular implications for how to construct the MCQs and distractors when aiming to identify commonly held misconceptions.

3.1.3.2 Extended response

It was rare to see examples of extended response. Crisp and Ward (2008) reported on the use of essay questions following the presentation of a real life scenario on screen. These extended responses would either 1) require marking by the tutor or 2) feedback was given by the tool by either providing model answers or the use of multiple-response question 'Which of the following points did you include?'

The use of extended response questions raises issues again about utilising the affordances of the technology; there are tools available that would automatically mark extended writing (e.g. Pearson's Intelligent Essay Assessor is used by classroom teachers as a learning aid.

The software gives students immediate feedback to improve their writing, which they can revise and resubmit). Work by Shermis (reported in the New York Times, 2012) compared human graders and software designed to score student essays, Shermis reported virtually identical levels of accuracy, with the software in some cases proving to be more reliable. But nonetheless such automated marking systems still raise suspicions that key aspects of writing, including style and appreciation of poetry may be lost.

3.1.3.3 Interactive response

E-assessment provides opportunities for the stimulus of a task to be interactive as well as the response to be interactive. This section refers to the rarer cases of where the response, not just the stimulus, requires interaction between the learner and the technology. Hughes (2006) reported on a trial of computer mediated mathematics questions for 11 and 14-year olds. A question was considered to require an 'interactive' response if it had these features:

1. There was an animated element to the question which was more than just an illustration, but with which the pupils needed to engage by moving or controlling some part of it

2. The interaction with the question changed the appearance of the animation and so gave immediate visual feedback. This means that the interaction is two-way, i.e. the pupil interacts with the question and the software allows some response to the pupil, usually in the form of some visual feedback and

3. The pupil had some control over the animation or objects in the question.

For example, one question aimed to assess the understanding of properties of an isosceles, right-angled triangle. Pupils dragged one vertex of a triangle to make it right-angled and isosceles. In a question assessing understanding of ratio, learners chose what size to make a grid in order to show a given ratio of red squares to grey squares.

Hughes found that it was with the interactive questions that there was most evidence of the affordances of the technology affecting pupil behaviour.

The NRich website provides resources for school mathematicians www.nich.ths.org; these include interactive tasks like those described in Hughes' work above which could be used for formative or self-assessment.

3.1.3.4 Implicit versus explicit feedback

Questions which require an interactive response typically give implicit feedback. Boyle's 2006 paper highlighted a distinction between two types of feedback

Extrinsic feedback is given following the completion of an activity and states whether the attempt was right or wrong. This type of feedback requires that the technology apply some algorithm to translate the student's action or response into feedback, e.g. by presenting a score or the appearance of a 'shiny star'.

Intrinsic feedback does not occur after the learner has submitted an answer, but is intrinsic to the act of working on the task. E.g. the question below (Hughes, 2006) gave immediate intrinsic feedback in response to the student's action; the change in the red shape provided

immediate visual feedback as to what shape the student had created. Then they would interpret this intrinsic feedback to decide if they had reached their goal of creating a pentagon with one line of symmetry. This type of intrinsic feedback allows the student to modify their behaviour and actions before submitting their answers, a strategy which is more akin with IBE than traditional approaches.

Rotate the tiles to make the red shape in the middle
a pentagon with one line of symmetry.

Click on any tile to rotate it.
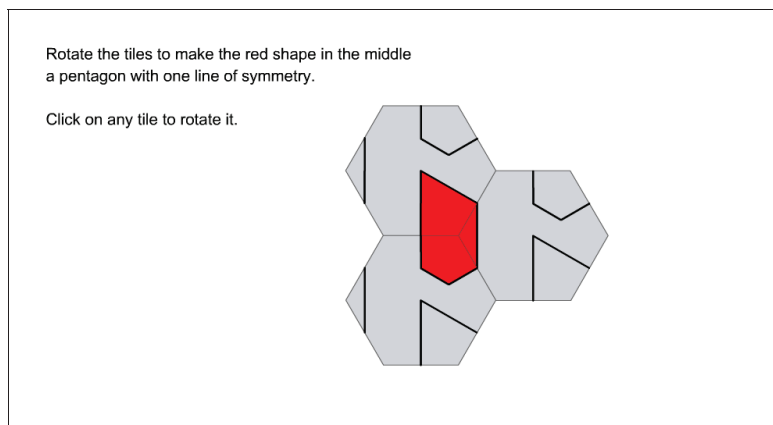
Similarly, the ARIES! Science online tool (Koenig, 2011) provided implicit feedback as the learner received feedback that was not in the form of a score or reward; learners received implicit feedback as they communicate with a student avatar and as they are party to conversations between two animated characters.

In Boyle (2006) we saw warnings that when designing feedback we need to be wary that e-feedback can lead to what was called 'superficial behaviour'. For example, in the mathematics example above, disengaged students were observed to be carrying out was could be called 'mindless clicking'; they were continually clicking on the hexagons receiving some kind of visual 'reward' for clicking on interactive objects, though not engaging with the goal of the task. Cook and Crabb (2002) asked how computer-based learning could be designed to maximise cognitive engagement and stimulate thinking rather than what they called 'random button pressing'.

Kennewell (2008) reflected that when teachers first adopted ICT as part of their practice, there was a tendency for interactivity to be superficial and authoritative, which the pentagon question above could be described as. Kennewell argued that it was only when technology was embedded in teachers' pedagogical knowledge did the technology contribute to deeper, more dialogic interaction amongst students.

3.1.3.5 E-portfolio

Broadfoot et al. (2013a) described e-portfolios as an information repository, a personal development record which provides a structure for the organisation of learning and collabora-

tion. An e-portfolio is a collection of digital objects showing evidence of a student's work. The technology provides an online system to manage the sharing of this work and to communicate feedback to students (Kimbell et al., 2009). When using e-portfolios, learners need personal online space for recording and evidencing attainment.

JISC (2007) proposed that e-portfolios were useful in promoting 21st C skills as when using e-portfolios, students are required to demonstrate skills of command of software, use web technologies and digital images, communicate electronically, solve problems and present and collaborate.

3.1.3.6 Other response types

Other response types seen or described in the literature included extended matching, repositioning objects by dragging and labelling diagrams.

3.1.3.7 Scratch pads

Onscreen scratch pads provide a means of catching learner's rough notes or sketches through touch screen technology. Learners are able to use either freehand 'writing' on-screen, and their device will capture the images and store them, or they can input via a keyboard. Hughes et al. (2011) reported on the development of a tool (at Pearson) named 'Overrite' which aimed to bring the experience of answering a mathematics question on-screen closer to the familiar experience of working on paper. Students could, via the keyboard and mouse, make jottings and annotations onto objects on the screen and save notes. This tool also aided assessment by collecting information on the processes students used to answer questions. One concern raised in relation to tools like this is that they are translating paper tasks to screen for the sake of comparability of the experience, rather than recognising that technology offers many affordances which can enhance learners' experiences, rather than trying to replicate what is done on paper (Kennewell, 2001).

3.1.4 Feedback

A longer section relating to the range of types of and audiences for feedback appears in section 3.8, but here we wish to highlight that the ease of provision of immediate and detailed feedback is one of the most valuable affordances of the technology (Kennewell, 2001).This supports the learner and can then lead to the provision of hints and appropriate learning activities, including Integrated Learning Systems.

3.1.5 Intervention type

One benefit of using technology is that learners' responses can be analysed and codified to provide not only useful feedback, but pointers to the students about how to progress. These pointers can include interventions which link the assessment outcome to the assessment objective, so truly supporting formative assessment.

Crisp and Ward (2008) provided learners with references to further reading and links to learning resources available on the web. Some other forms of help included the option to read a short explanation of relevant ideas before moving on from the stimulus to the assessment questions. Other types and examples of interventions are described in section 3.2.1 below.

3.2   Assessment/teaching link

The link between assessment and teaching is key, and central to the process of formative assessment, and this also applies to the relationship between e-assessment and e-learning.

Bennett (2002) stated that technology would change assessments dramatically. He proposed that electronic test development would evolve through three generations.

He described the then current (2002) generation of CAAs (computer Assisted Assessments) as 'migrational', whereby on-screen tests were simply a migration of existing paper tests onto the screen without reconceptualising the process or the content. This migration of tests, he argued, failed to 'realise the dramatic improvements that the innovations could allow' (Bennett, 2002). As such he argued that these first generation tests didn't utilise the functions of the technology to change the test for the better.

He predicted that the second generation of CAA would exploit the features of the technology, for example, by the use of colour, sound, animations, video and the integration of interaction between the test taker and the test.

Bennett predicted, finally, that 'Generation R' tests would evolve (the R standing for 'reinvent'). Generation R tests will be assessments so closely integrated into teaching and learning, that they will be indistinguishable from learning materials.

Formative assessment requires that the delineation of assessment and teaching is blurred, with assessment being a subset of teaching, suggesting that Bennett's 'Generation R' assessments would be akin to formative assessment.

Bennett's conceptualisation of Generation R assessments also brings up the issue of the compatibility of e-assessment with traditional teaching/learning methods; if technology is used to support teaching and learning, it follows that to ensure that an assessment is valid, it also needs to be supported by the use of technology.

Kennewell (2008) recognised that some affordances of the technology relate to administrative rather than pedagogic concerns. However, in many CAA systems, Thomson Prometric reported in 2006, a complete electronic alternative to the existing assessment process is provided; the process can include the activities which can be more efficiently administered electronically than in paper-based systems:

1. Registering students and storing their details

2. Authoring questions

3. Pre-testing questions

4. Storing questions and the associated data

5. Delivering test to students

6. Delivering tests to markers

7. Marking

8. Storing completed papers

9. Storing outcomes

10. Converting outcomes into useful feedback (whether that be a pass/fail, mark, grade or report).

These may be benefits for administrators, but few would be recognised as benefits to learners. Learners benefit when assessment and teaching/learning are connected and relate to the same constructs. To be valid, an assessment must exploit the affordances that technology can bring to learners, not just administrators.

Neumann (2010) described the relationship between the use of technology in teaching and learning in Higher Education (HE) statistics as close. His work on using e-assessment with statistics students was possible because, in HE statistics courses, technology is central to teaching and learning, and hence can be more validly used in assessments and with less controversy than at lower levels of education.

Halperna et al. (2012) described the evidence-based design process behind a secondary school science e-learning programme with integrated assessment, 'Operation ARA' (previously Operation ARIES!). Developers identified good practice in e-learning and then integrated assessment into that programme.

Kennewell (2008) showed a specific focus on interactivity; his concern was for linking the concepts of interactive teaching and interactive technology, consequently championing interactive e-assessment. He argued that a shifting balance in the classroom towards dialogic would bring improvements to the learning process. His argument was that the nature of interactivity was more influential than the more general use of ICT, the latter of which could simply relate to administrative benefits of the technology. JISC (2007) described how effective practice with e-assessment involves the linking of assessment and learning, with content available online via a learning platform and the contact time (lectures) between teachers and students being used to refine understanding rather than introduce a topic. This would require formative and/or diagnostic assessment to be used to identify what understanding to address during that contact time and at what level. In this example, the impact of technology on learning would be to increase the requirements for formative assessment.

3.2.1 Examples of interventions

Intervention is central to formative assessment; teachers use outcomes of formative assessment to select or devise interventions for learners to further their learning. Technology allows the process of translation of student response into an appropriate intervention to be speedy and based on evidence and good pedagogic principles. Broadfoot (2013d) defined learning analytics (further described in section 3.8.1.1) as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs. Learning analytics are generally algorithms used by the technology resulting in decisions or suggestions about which type of interventions should be adopted.

3.2.1.1 Intelligent tutoring

Intelligent tutoring can refer to a huge variety of types of e-learning, including a simple clue or hint, the addition of scaffolding, feedback specific to particular responses, links to learning resources including animations, references to further reading or interactions with virtual or real tutors or peers.

Crisp and Ward (2008) incorporated in their feedback to students, references to further reading and links to learning resources available on the web. Other forms of help included the option to read a short explanation of relevant ideas before moving on.

Feng et al. (in Pellegrino and Quellmalz, 2010) described the 'ASSISTment' system, which is a pseudo-tutor for middle school level mathematics. The system uses scaffolding questions, optional hints, and buggy messages (specific feedback given after student errors) for each item. Students must eventually reach the correct answer, and scaffolds/hints are limited to avoid giving away the answer.

Halperna et al. (2012) Operation ARA (previously ARIES!) described how students move through the interactive chapters of the system, they receive computer-generated tutoring that varies depending on how well the student responds. The type of tutoring that students receive following each chapter is determined by the number of questions they answer correctly about the chapter they have read.

Implications for the design of these types of interventions are that e-tutors need to both correctly gauge and adapt to the student's current level of understanding. A successful adaptive tutor chooses problems that specifically address the level of the student's prior knowledge and take previous test scores into consideration. In order to maintain engagement during vicarious learning (where the learner is party to a discussion between a virtual student and a virtual tutor or two virtual students) the learner is asked to respond to questions about the tutoring situation. For example, the virtual teacher might ask the human student whether the virtual student understands the concept or whether the virtual student's answer was correct.

3.2.1.2 Computer Adaptive Testing (CAT)

Quellmalz (2009) defined CAT as procedures in which items are selected based on the examinee's prior response history and an underlying model of proficiency, developed to reduce time testing and examinee burden. Such adaptive testing can integrate diagnosis of errors with student and teacher feedback. It can tie assessment more closely to process and contexts of learning and instruction.

In its simplest form, CAT automates decisions about which questions to present in a sequence of questions as determined by the performance on the previous question. The complexity of the algorithm used to analyse previous performance can vary considerably.

DreamBox Learning' (available at www.dreambox.com) differentiates instruction and adjusts difficulty levels and the number and type of hints to give learners, based on tracking pupils' responses to several different questions. Many sources of evidence from many tasks relating to both processes and content can be used to input to decisions about what task to present to a student next. Quellmalz (2009) described how the computer's ability to capture pupils' inputs permits collection of evidence of problem-solving sequences and strategy use

ASSIST

as reflected by information selected, numbers of attempts and time allocation. These can be combined using statistical and measurement algorithms for patterns associated with varying levels of expertise and then students can be directed as appropriate. This information can also be relevant for assessing against competences.

Consideration needs to be made of which data it is useful to capture for the learner and for the teacher. The effective use of CAT requires reconceptualising assessment design and its use involves tying assessment more directly to the processes and contexts of learning and instruction.

Two concerns relating to CAT arose from the literature:

1. Use of item banks to randomly select questions to present to learners was seen by students as unfair (Voelkel, 2013). Voelkel found that different variants of computer marked questions could behave differently in terms of their level of demand and the construct being assessed.

2. Pachler et al. (2009) found widely differing theoretical emphases being applied in e-assessment development in the literature, as well as varying views of 'adaptivity' as a core component of e-assessment processes.

3.2.2 Re-submitting issues

Whether students are permitted to change their answers in an e-assessment usually depends on whether the assessment is high stakes or low stakes and whether it is formative or summative.

Neumann (2010) reported on a statistics simulation used for summative assessment in HE. Feedback was minimal because of the summative purpose of the assessment and the fact that students could attempt questions more than once. Voelkel (2013) described how feedback was tailored by providing more and different feedback after second or multiple attempts at the same questions. Scholar (www.scholar.hw.ac.uk) used a method of being able to re-submit an answer before the correct answer is revealed, this allows a more iterative process to understanding. A number of repeated demonstrations, for example, in mathematics, allow different pupils to vary the amount of reinforcement received depending on their confidence and/or understanding.

This section brings up implications for giving feedback (or not) between attempts at a task. In high stakes assessments this can still provide useful feedback to the teacher, even if this is not shared with the student.

3.3   Inquiry based education

The ASSIST-ME proposal described Inquiry Based Education (IBE) as an umbrella term, encompassing a wide range of teaching approaches that can enhance student motivation which have the potential for enhancing learning outcomes.

Inquiry-based STM education includes students' involvement in questioning, reasoning, searching for relevant documents, observing, conjecturing, data gathering and interpreting, investigative practical work and collaborative discussions, and working with problems from and applicable to real-life contexts (Anderson, 2002). Inquiry-based STM-education is not a

new teaching method, but it is often used as a contrast to more traditional teaching approaches, such as those where the teacher presents results and methods which the students are then trained to apply. Giving students an active part in learning is in accordance with many teachers seeing the pedagogical principles of constructivism as the foundation for understanding and implementing inquiry-based learning (Llewellyn, 2007).

Only a few of the articles found here, where e-assessment and e-learning are discussed, explicitly reported on the assessment of IBE.

Neumann (2010) assessed functional knowledge using MCQ questions following a simulation relevant to HE statistics and Feldman and Capobianco (2008) reported on the use of an electronic voting system in which assessment items were designed to be 'consistent with constructivist and active-learning pedagogies'. Feldman and Capobianco highlighted that, for teachers to successfully implement a conceptual learning approach (i.e. IBL) to physics teaching, they may need to make significant changes to their teaching methods. Scholar (www.scholar.hw.ac.uk) a secondary science and mathematics environment, including simulations, utilises animated graphs which could be used as stimuli for IBE.

Of the examples of assessment of IBE that we found, we have selected examples of e-learning environments that were impressive in their scope and methods, for example in science Operation ARIES! (Koenig 2011) and SimScientists (Pellegrino and Quellmalz, 2010) were both e-learning environments which integrated the learning and assessment of IBE.

As described earlier, Operation ARIES! (Koenig 2011) includes an assessment model in which learners are presented with inaccurate science information through the medium of newspaper headlines and television news channels and must ask questions to ascertain the truth. Students engage in solving a problem through dialogue interactivity whereby students learn by engaging in conversations and tutor groups. Halperna (2012) reported that students repeatedly practiced and applied concepts in different contexts and from different domains within science, and argued that this variability enhanced transfer of knowledge and skills.

Halperna (2012) also discussed, relating to competences and communication, students answering a number of multiple choice questions at the end of each chapter of an e-book and then participating in trialog discussions with avatars where the understanding of the material from the chapter is clarified and reinforced.

Pellegrino and Quellmalz's (2010) work on the SimScientists software illustrated ways that assessment tasks can take advantage of the affordances of simulations to represent challenging inquiry tasks, indeed many e-learning or e-assessment tools, which have claimed or attempted to assess IBE, have used simulation. In order to assess IBE there must be observable evidence of IBE, so assessments must provide opportunities for learners to engage in IBE. Pellegrino and Quellmalz (2010) described two national projects in which the use of interactive simulation tasks enabled the assessment of IBE:

1. The 2009 NAEP science framework and specifications drew upon science simulations work (reported in Wylie and Dolan, 2013) in developing their rationale for the design and pilot testing of interactive computer tasks to test students' ability to en-

ASSIST

gage in inquiry practices. Such innovative items were included in the 2009 NAEP science administration.

2. Wylie and Dolan (2013) also reported that the new 2014 Technology and Engineering Literacy Framework for NAEP will be entirely computer administered and will include specifications for interactive, simulation-based tasks involving problem solving, communication, and collaboration related to technology and society, design and systems, and information communications technology.

Similarity, the 2015 PISA framework for scientific literacy (OECD, 2013, available at www.oecd.org/pisa) considers the possibility of assessing collaborative science problem solving skills by computer, in a summative high stakes assessment.

3.4   Competency-based learning

The ASSIST-ME proposal understands competence to mean a combination of skills, knowledge, characteristics, and traits that contribute to performances in particular domains. There is not a universal agreement on the terminology of competence. In this project we will use the word competence for both a competence, referring to the concept in general and a level of ability, and a competency, referring to a particular demand that a person may or may not be able to meet, and the plural form competences, to reflect an integration of understanding and attitude into the concept.

The more complex the learning goals, the more difficult they are to measure. The understanding of competences as the ability to cope with relatively complex challenges in everyday life means that assessment methods necessarily have to be relatively advanced, flexible and process oriented. This suggests that in order to assess IBE related competences, there is a need for learners to engage in IBE as part of their assessment.

Operation ARIES! (Koenig, 2011) is designed to assess and teach critical thinking about science, a core competency. It uses intelligent tutoring and makes use of animated characters to enable students to develop and exhibit relevant competences, including inter personal skills. The way in which Operation ARIES! requires that students generate their own questions about abbreviated research descriptions in order to determine whether the research is flawed in also in line with CBL.

3.5   Formative assessment

The ASSIST-ME proposal describes formative and summative assessments as similar in that they involve the collection, interpretation and use of data for some purpose. They are mainly identified and distinguished from each other by the purpose of the assessment, but often also in the way data is collected. Formative assessment has the purpose of assisting learning and for that reason is also called 'assessment for learning'. It involves processes of 'seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning and where they need to go and how best to get there' (Assessment Reform Group, 2002).

Under this theme of formative assessment, some key areas were identified in the literature which have relevance for the design of ASSIST-ME e-assessments:

- Self-assessment, incorporating confidence ratings

- Peer assessment, relating to both the assessment of collaborative and individual work

- Diagnostic assessment, including in relation to the diagnosis of misconceptions

- The iterative, personalised nature of some adaptive assessments and

- The interaction between and roles of the teacher, learner and computer, including reference to how e-assessment can support discussion.

Note the omission of feedback in the above list: this is an extremely important aspect of formative assessment. Whitelock (2006) identified one driver for the implementation of e-assessment being improving learning through faster feedback which he related to increases in student retention, flexibility, support in coping with large student numbers, providing objectivity in marking, the more effective use of VLEs (Virtual Learning Environments), and more reflective learners who are more in control of their learning. Feedback is so important that rather than make it a subheading of formative assessment, it is discussed in section 3.8 which includes all feedback-related findings.

3.5.1 Self-assessment

Pachler et al. (2009) wrote that learner self-regulation was a core feature in assessment and is linked to motivation and emotional factors which affect learners' engagement with feedback.

Supporting students in judging their own learning or performance can help develop the skills of self-regulation. Pellegrino and Quellmalz (2010), reporting on the interactive game Sim-Scientists, illustrated ways that assessment tasks could take advantage of the affordances of simulations to represent generalisable, progressively complex models of science systems which promote metacognitive skills through self-assessment.

3.5.1.1 Confidence ratings

A number of assessments found in the literature required that learners complete confidence ratings, in which they rate how confident they are in their answer.

JISC (2007) proposed that confidence based marking could promote a deeper level of learning by challenging learners to evaluate certainty in their answers so that they could address gaps that they discovered in their knowledge. A learner's confidence is affected by their self-efficacy, which is one's belief on one's ability to succeed in specific situations. One's sense of self efficacy can play a major role in how one approaches tasks, however the danger is, while students with a strong sense of self efficacy are more likely to challenge themselves with difficult tasks and be intrinsically motivated, students with low self efficacy believe that they cannot be successful and are thus less likely to make an extended effort and may consider challenging tasks as threats to be avoided.

Crisp and Ward (2008) also captured confidence ratings by asking users to indicate their confidence after answering each automatically scored question. At the end of an assessment an analysis of the student's metacognition was reported, based on the given ratings for use in formative assessment. The ratings described by Crisp and Ward (2008) didn't influ-

ence a student's path through an assessment or the immediate feedback given, whereas Swithenby (2006) reported on how pupils were given options depending on how they rated their confidence in their answer at the point when they submitted their answer. These options included: submit; hints; show answer; review part; display mathematics; and give clues.

3.5.2 Peer assessment

Broadfoot et al. (2013c) stated that successful peer assessment required individual responsibility from students, interdependence on peers, and trust within the group. Practitioners should recognise that students can be anxious about the ability of their peers to assess learning, their own abilities to assess others' work and the overall validity of peer assessment.

Examples of peer assessment in the literature included

- Group peer assessment – in which each member of a group who had collaborated on a task judged each other's contributions and

- Individual peer assessment – in which a piece of work/performance/response carried out by one student was evaluated by one or more peers.

3.5.2.1 Group peer assessment

Electronic voting systems can be used to gather group peer assessments, for example, using an electronic voting system to evaluate previous students' practical work against specified marking criteria.

Digital technologies have the potential to support collaborative learning and assessment practices, such as undertaking knowledge building activities, co-evaluation and social interaction. A case study (Broadfoot et al., 2013c) using computer-supported collaborative learning (CSCL) limited the participation of students in some assignments and generally low-quality assessment reports.

Kennewell (2008) reported on software for group work incorporating features to capture contributions from different students. An example activity was to develop a concept map for photosynthesis.

Broadfoot et al. (2013b) described 'crowd-sourced grading' involving weekly peer evaluation of student blogs. McKinsey and Company (2013) described a process of evolving better answers through collaboration: Students were set assignments to write blogs for sharing test results of their designs and receiving comments from professors and classmates. Some argue (in Broadfoot et al. 2013c) that using tools like wikis or blogs in group assessments can further exclude some students by benefitting those who are already users of social media.

3.5.2.2 Individual peer assessment

A simulation of individual peer assessment is found within the Operation ARIES! science environment (Koenig 2011) where students receive feedback and tutorage from two characters from the program throughout. Two characters have conversations with each other in the presence of the learner, using virtual peer tutoring.

3.5.3   Diagnostic Assessment

The availability of measurable, detailed descriptions of the constructs and factors to be assessed is the essential prerequisite for the construction of diagnostic items, as well as the tests. Developing an online diagnostic assessment system for grades 1 to 6 CRLI (2009) described constructs being assessed in terms of misconceptions that students may have about the subject.

3.5.3.1 Misconceptions

We found examples of how technology can provide evidence of learners' misconceptions and hence, use this to enable them to progress (Wylie and Dolan, 2013). Feedback to teachers can aid clarification of which misconceptions are held by students (Voelkel, 2013).

Furse (2009) described it as straightforward to incorporate misconception handling into an e-assessment system if the author knows of suitable misconceptions. Wylie and Ciofalo (2008) reported on the creation of a bank of items for High School mathematics and science teachers that drew on the misconception literature. These high school formative e-assessments were through multiple choice questions and each multiple-choice item that was developed drew on at least one previously identified student misconception. These question types have different implications for item development than those which are not misconception based (Wylie and Dolan 2013).

3.5.3.2 Progression levels

Also reported in Wyllie and Dolan (2013) was a middle school mathematics project which focussed on the progression between levels of understanding, rather than just on categorising the learner according to in which level they sat. The project used two kinds of assessments:

locator assessment: computer delivered and places student within three learning progressions and

incremental tasks: which explicitly target a transition between levels, rather than the levels themselves.

3.5.4   Teacher/learner/computer relationship

Building on research principles of the Assessment Reform Group which firmly put an emphasis on Assessment for Learning and the relationship between the teacher and the student, the e-assessment association of the UK (Beevers et al., 2011) believe that software solutions designed for formative assessment should also follow the ten ARG principles:

1. be part of the effective planning of teaching and learning

2. focus on how students learn

3. be able to be central to classroom practice

4. promote professional skills for teachers

5. be sensitive and constructive, being aware of emotional impact

6. foster learner motivation

7. promote commitment to learning goals and assessment criteria

8. help learners to know how to improve

9. develop the learner's capacity for self-assessment and

10. recognise a range of educational achievement.

Feldman and Capobianco (2008) used whole class responses using an electronic voting system to prompt discussion where they reminded us that formative assessment was supported by classroom discussion and the technology in their study-aided classroom discussion.

To facilitate one to one discussions between learners and teachers there needs to be focus on the relationship between the teacher and the learner; Wylie and Dolan (2013) placed this relationship at the centre of formative assessment. They warned that the use of external 'off the shelf' tools could distance this relationship: 'To carry out formative assessment, teachers must be proficient in developing their own evaluative tools as part of their instructional practice.' (Wylie and Dolan 2013 p1)

McKinsey and Company (2013) extended these relationships to include parents giving the example of Ultranet, a student-centred learning environment which allows students, teachers and parents to connect and collaborate to improve learning outcomes.

3.5.4.1 Prompts for discussion

Wylie and Dolan (2013) viewed the output of some formative assessment as a stimulus for teacher-led discussion. This requires that items are written to stimulate discussion, not to summatively assess. Discussion could be one to one, whole class, small groups etc. They stress that the tool collects the evidence and the responsibility is with the teacher to use the evidence appropriately.

The scenario-based assessments described by Crisp and Ward (2008) could be used to stimulate discussion amongst groups of learners, but were also intended to be usable without the guidance of a facilitator. In the assessments, the computer provided some of the guidance, probing and directing mimicking the role that a discussion leader would normally provide in case methods.

3.6   Summative

Summative assessment has the purpose of summarising and reporting learning at a particular time and for that reason is also called 'assessment of learning'. It involves processes of summing up by reviewing learning over a period of time or checking-up by testing learning at a particular time.

The SimScientists game provides teachers with feedback on student and class progress both on general summative measures (e.g. time to completion, percentage correct) and on more specific knowledge components (Pellegrino and Quellmalz, 2010).

Concerns about using e-assessment for high stakes summative purposes include collusion, plagiarism, recognition of partial achievement, logistical problems of simultaneously allowing access to computers for a whole class (Swithenby, 2006) and security concerns (Dennick 2009). Summative assessment of collaborative work had added complications; Swithenby (2006) claimed that for group work it was easy to monitor the amount of time pupils spent on an activity or contribution, but difficult to judge the quality of it. One proposed solution was to use more peer assessment.

To mitigate against the risks of using e-assessment for summative purposes, Swithenby reported on how Warburton (2006) suggested a gradual, low risk strategy through quizzes and progress checks leading to one summative assessment (high stakes). Another means of using formative and summative assessments in the same platform was the repeated used of frequent formative assessments followed by a final summative assessment.

3.7   Formative / summative assessment link

There are a number of affordances offered by the technology (Kennewell, 2001) some of which relate to administration (e.g. reducing postal traffic, speeding up marking, etc.) and some to pedagogy (e.g. interactivity, simulation, giving students control, providing environments/experiences not possible on paper, giving detailed and immediate feedback). Administrative benefits of e-assessment have been utilised for high stakes summative assessments (e.g. the electronic marking of examination scripts), but there has been less take-up of the pedagogic affordances of technology for summative assessment.

As Boyle (2006) pointed out, there is a conflict when considering the use of e-assessment for high stakes summative assessment, because a high stakes environment is not conducive to innovation or risk taking. But he predicted that formative assessment would be the vehicle for innovation as there are more opportunities for risk taking in a formative assessment context.

3.7.1   E-assessment as a support for linking formative and summative assessment

E-assessment can support the link between formative and summative assessment. Pachler et al. (2009) report that within e-assessment there is a tendency to conflate formative and summative assessment. VLEs bring together learning and assessment and consequently formative and summative assessment. Bennett's 2002 vision of 'Generation R' e-assessment included assessments that are so closely integrated into teaching and learning, that they will be indistinguishable from learning materials. Broadfoot et al. (2013a) describe how making assessment and instruction simultaneous would support the integration of formative and summative assessment.

Broadfoot et al. (2013b) propose that integrating formative and summative assessment would be more meaningful for students; using an integrated assessment, learners could benefit from regular feedback which supports learning. They argue that the link could also contribute to an overall picture of learning, is more authentic, has the potential to track progress, aggregate data, create multi-media platforms for feedback and review, accumulate evidence and help learners understand the connections between learning and assessment.

3.7.2   Digital objects for both formative and summative assessment

There are many electronic tasks which can be effective as learning objects for formative assessment or for summative assessment.   Neumann (2010) used simulations that were designed for use in summative assessment, but would be appropriate for use in formative assessment or for teaching and learning.

The evaluation of the e-portfolio system E-scape (Kimbell et al., 2009) showed that the e-portfolio can be used for formative assessment as all student activity is recorded and for summative assessment in which case students can edit and select which work to submit for judgement. This supported the argument that it is what is done with the information that determines whether it is formative or summative assessment, and the same tasks could be used for either.

But some would argue that formative and summative assessments are different. For example, Voelkel (2012) argued for the separation of summative and formative processes, based on a view that the use of the same assessment for formative and for summative purposes is not always beneficial for learning. Perhaps this depends on the type and quality of the assessment. Some examples of formative e-assessment can be argued to be serial summative assessment (e.g. Pachler et al. 2009). Formative assessment appears to be equated with 'low stakes' assessment, or 'practice' assessment in preparation or contributing towards high stakes summative outcomes. This does not necessarily reflect the principles of formative assessment as set out by the Assessment Reform Group (shown in section 3.5.4).

SimScientists is another example of an environment which uses assessments for both formative and summative purposes. The SimScientists (Quellmalz et al., 2012) includes assessments designed to supplement state science test evidence by providing science assessments that are

1. embedded within curriculum units that could serve formative assessment purposes by providing immediate feedback, monitoring progress, and informing needed adjustments to instruction and

2. administered at the end of a unit as summative measures of proficiency on the targeted science content and inquiry practices.

3.7.3   Games as a means of blurring the formative /summative boundaries

Games use actual learning as their basis for assessment: their assessments are built on problem solving and facing challenges (Gee and Schaffer, 2010). Games:

- assess whether a player is ready for future challenge

- track information over time

- are designed in levels, and each level requires that students have mastered the previous level and that they learn new skills on the new level.

These three characteristics are compatible with both formative and summative assessment.

3.8 Feedback

One advantage of using e-assessment is the ability of the technology to provide quick and detailed feedback (Kennewell, 2003). Feedback supports the learner and can then lead to the provision of hints and appropriate learning activities.

It is clear from many sources that feedback is most effective when it is instant, differentiated and individualised (e.g. Swithenby, 2006, Pellegrino and Quellmalz, 2010).

A number of papers (e.g. Neumann, 2010, CALM in Beevers, 2011), Scholar, www.scholar.hw.ac.uk, and Voelkel, 2013) cited the impetus for using e-assessment as the opportunities that it brought for giving meaningful feedback to large numbers of students.

Boyle (2006) advised that in designing feedback we need to be aware of the audience for the feedback. He raised the questions: Is the audience the teacher and/or the student; Will one type of feedback be appropriate for both audiences?

The discussion of issues arising relating to feedback is most usefully structured by audience: section 3.8.1 describes what was found in relation to feedback that is provided to the teacher and section 3.8.2 describes findings in relation to feedback as it is given to learners.

3.8.1 Output to teacher

Key decisions need to be made in the assessment design process relating to what feedback the teacher would benefit from receiving. E-assessment generates rapid, reliable data on learners' progress and can indicate which learners are at risk and provide prompts for remedial action (JISC, 2007). Through the CALM project, for example, (Beevers et al., 2011) lessons were learned including which details to record for the reporting process.

With such a variety and depth of data available, these decisions are not simple. Data goes way beyond just scores and performance data and can relate to processes and contexts in which assessments were tackled, for example the ALTA system (Adaptive Learning Teaching Assessment) (2009) includes the possibility to collect both pupil and class data and to collect longitudinal data and trends.

3.8.1.1 Learning analytics

Learning analytics use data about learners to optimise learning. Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs. It interrogates learner-based data interactions; techniques include predictive modelling, user profiling, adaptive learning and social network analysis (Broadfoot et al., 2013d).

Learning analytics can be used as a precursor to intelligent tutoring (Section 3.2.1.1) during which data is analysed to establish which type of learning or support should be tackled next. For example, McKinsey (2013) describes how through using mobile technology for learning and assessment, learning can be adapted to individual pupil's understanding and pace. This recognises the need to analyse patterns of pupil behaviour (not just performance) over a number of assessments. Unlike summative assessments which provide a snapshot of performance level, this enhanced type of feedback can be generated using learning analytics from numerous assessments over time.

3.8.1.2 Feedback level

Reports to teachers can be provided at a number of levels including, for example, individual learner, group, class, year, whole school or cohort. Boyle (2006) concluded that formative reports should be threefold: one for the student audience, one for the teacher audience reporting individual student feedback and a third for teachers reporting on the whole class or a group of students. Item-level data is also of use for teachers in the assessment of learning.

Teacher reports are commonly presented via a dashboard (Broadfoot et al. 2013d) in which teachers can switch between individual, class or year group views. The E-scape portfolio system (Kimbell et al., 2009) used a timeline onto which all digital objects and student work were placed. This allowed both the teacher and student to see a process of learning and thinking, which served as an effective reflective tool for the student and evidence of process for the teacher making assessment judgements.

3.8.2   Feedback to student

Gipps (2005 in Crisp and Ward, 2008) described the facilitation of formative feedback to learners as one of the key reasons for the growing use of e-assessment in UK higher education, but stressed that only assessments providing adequate feedback would enhance learning. Gipps commented that 'the developments in automated, diagnostic feedback in short answer and multiple-choice tests are . . . potentially very valuable. If feedback from assessment could be automated, while maintaining quality in assessment, it could certainly be a powerful learning tool' (p. 175).

While dashboards can be used for teacher feedback, Broadfoot (2013d) also described how dashboard systems allow learners to monitor their own academic or behavioural activity, to access relevant strategies and support and to compare their performance to previous students/classes. However there are a number of challenges and ethical debates and concerns (e.g. demotivation of less able) to be considered here.

Gibbs and Dunbar Goddet (2007 in Voelkel, 2013) found that 'giving out clear goals and standards had little effect on learning, and that it was much more helpful when students received plenty of feedback' (Voelkel, 2013).

3.8.2.1 Characteristics of feedback

Shute (2008, in van der Kleij et al. 2012) suggested making a distinction between feedback type and feedback timing. Students value high quality, actionable feedback (Black and Wiliam, 2009).

The themes of type, timing and level of detail of feedback are developed below, along with some findings relating to how students action the feedback that they receive.

3.8.2.2 Type of feedback

Hattie and Timperley (2007 in van der Kleij et al., 2012), in their analysis of feedback in computer-based assessment for learning, distinguished four levels at which feedback could be aimed, which is an expansion of a previously developed model by Kluger and DeNisi (1996, in van der Kleij et al., 2012). The levels distinguished are the self, task, process, and regulation levels.

- Feedback at the self level is not related to the task performed but is aimed at characteristics of the learner. Praise is an example of feedback at the self level. Feedback at the self level is not seen as effective for learning because it does not provide the student with information regarding how to achieve the intended learning goals.

- Feedback at the task level is mainly intended to correct work and is focussed at a surface level of learning (e.g. knowledge or recognition); for example, the student is told whether the answer is correct or incorrect.

- Feedback at the process level relates to the process that was followed in order to finish the task. In this case, for example, a worked-out example is given.

- Feedback at the regulation level is related to processes in the mind of the learner, like self-assessment and willingness to receive feedback. In the ideal situation, the feedback is adapted to the current level of the learner.

Hattie and Timperley favoured feedback aimed at the process or regulation level in order to enhance learning. Van de Kleij et al. (2012) gave clear indications as to the type of feedback that learners perceived to be most useful for learning; learners' opinions gathered via questionnaires indicated that learners perceived immediate and delayed feedback to be more useful for learning than delayed knowledge of results only.

3.8.2.3 Timing of feedback

The timings with which the feedback from e-assessments is delivered to learners can vary and this variation can have an important effect on how useful it is to the learners. Boyle (2006) discussed how e-assessment designers need to be wary that feedback is given at the right time whether this be at the end of a question, a series of questions or a session.

Others differentiate differently timed feedback as 'immediate feedback' being feedback given immediately after completion of an item and 'delayed feedback' as feedback given directly after completion of all of the items in an assessment. Shute (2008 in Van der Kleij et al., 2012) also attempted to distinguish immediate and delayed feedback by claiming that immediate feedback is (usually) provided immediately after answering each item while the definition of 'delayed' is more difficult to make, since the degree of delay can vary. In some cases, the feedback is delayed until a block of items has been completed. Delayed feedback could also mean feedback being provided after the student has completed the entire assessment. However, feedback can be provided an entire day after completion of the assessment or even later. The nature of the assessment in terms of its summative or formative nature could affect which is more appropriate.

Van der Kleij et al. (2012) suggested that learners prefer immediate feedback to delayed feedback. A game, quiz or a simulation can give instant feedback. In these situations, immediate feedback is important as the task is acting as formative or self-assessment, for example, with the Scholar programme. Voelkel (2013) looked at science e-assessment in HE. Feedback was not given directly after a whole test, rather it was given at a lecture a week later; less than 60% of students reported that they found this delayed feedback useful. Beevers et al. (2011) also highlighted the immediacy of feedback as an ingredient for successful formative e-assessment.

3.8.2.4 Frequency of feedback

Related to the timing of the feedback is the frequency of the feedback. Quellmalz et al. (2012) highlighted the comparisons between frequency and ownership of formative assessment, whereby formative assessment went beyond annual, high-stakes tests to multiple assessments over time and in time for teachers to tailor instruction.

Halperna et al. (2012) commenting on Operation ARIES! described how, as learners progressed through, their understanding of the concepts and principles of science were repeatedly tested. This required that learners demonstrated their learning consistently throughout the programme. Through this active engagement it was felt that more deep learning was achieved as the learners were actively engaged with the material. Through this model, Halperna et al. (2012) felt that the learners could become judges of their own performance and rely less and less on external knowledge of results than they would with constant feedback. Using this principle in the design of Operation ARIES!, as learners progressed through the programme, they received feedback that was increasingly less frequent and less detailed.

3.8.2.5 Level of detail of feedback

Scores alone do not provide the necessary information for learners to use them as effective feedback (Gipps, 2005 in Voelkel, 2013). Some believe that, for learners, knowing which answers were correct is just as important as knowing which answers were incorrect (Hattie and Timperley, 2007 in Voelkel, 2013). JISC (2007) reported how online mock tests were useful when immediate feedback was given for correct and incorrect answers.

However, generally it is agreed that while assessments that provide grades and scores tend to increase the tendency for learners to adopt performance, rather than mastery goals, these grades and scores can increase motivation in the short term. In the longer term, however, the effect appears to be detrimental to formative processes and to learning (Black and Wiliam, 1998 in Pachler et al., 2009). So for e-assessments, it should be that more detailed feedback than scores alone are provided for the learner.

Hattie and Timperley (2007, in Voelkel, S., 2013) emphasised that effective feedback needed to provide information that specifically related to the task so that learners could develop self-regulation and error detection strategies and use the feedback to then tackle more challenging tasks. Furse (2009) felt that learners needed to have an explanation of the answer. It was not sufficient just to be told the correct answer especially as 'rushers' were likely to just want to get on to the next question rather than digest the answer. One solution was to leave the correct answer and its explanation up for 5 seconds before learners could proceed. A more sophisticated approach would be to use animated text. It is probably desirable for an e-system to also check whether the learners think they understand their mistakes.

3.8.2.6 Learners' use of feedback

This section arose from literature which showed a concern for how students use feedback. Questions were raised such as: Do learners use feedback?; What is the value of feedback and how can that be optimised?;   How can we encourage students to use feedback effectively? What support or advice do they need to use feedback optimally?

Halperna et al. (2012) suggested that feedback was important in that it provided information to the learner about his or her own performance, but that the learner still had to derive meaning from it. It may be that the way learners interpret feedback is what determines when it will be beneficial. Beevers et al. (2011) reported that feedback from e-assessment encouraged learners to take responsibility for their own learning. This opportunity for learners to use their feedback for ipsative assessment essentially encouraged learners to become proactive, self-critical learners rather than just using feedback for normative processes.

One key difference with the immediate feedback offered by e-assessments is that learners can act on it 'there and then'. Swithenby (2006), and Jordan (2009, in Voelkel, 2013) found that interactive computer-assisted assessments allowed learners multiple attempts and with built-in feedback could engage learners in meaningful learning activities, as they were asked to act on it immediately.

3.9   Effects of e-assessment on the learner

This section discusses the impact that assessment can have on the learner. JISC (2007), Whitelock (2006) and Broadfoot et al. (2013) described some of the potential, positive effects of e-assessment on the learner:

- increasing the range of what is tested

- encouraging deeper learning

- fostering more effective learning for a wider diversity of learners

- presenting challenging yet stimulating ways to demonstrate understanding and skills

- more authentic experiences being offered, for example, through using simulations

- good quality, timely feedback

- linking to appropriate resources

- feedback including opportunities for further learning

- supporting personalisation: learners can progress at a pace and in a way appropriate to them, for example, e-portfolios helping learners to present themselves and their work in a more personalised manner

- allowing learners to realise their own potential

- on-demand summative assessments increasing motivation

- learners taking tests voluntarily if they are available anytime, anywhere which can in turn help to establish more regular patterns of study; learners have been more likely to test themselves more regularly than with pen and paper tests.

In the literature reviewed, three key effects that e-assessment could have on the learner were identified: motivation, learning gains and learner confidence.

3.9.1 Motivation

E-scape (Kimbell et el., 2009) showed how technology supported learning in itself was a significant motivator for 14-19 year olds and was in fact one of key tools in accomplishing higher levels of engagement and achievement with this age range. Broadfoot et al. (2013a) reported that the tools that could be used to support e-assessment, for example, wikis, blogs, social networking activities, podcasting and e-portfolios, provided richer activities that lead to improved learner engagement.

Crisp & Ward (2008) found evidence that e-assessment increased motivation and performance. They also felt that we should make use of the motivational benefits of e-assessments by developing engaging, interactive formative assessments, which could be used either independently without teacher intervention, or in preparation for a classroom discussion or activity. Beevers et al. (2011) described the ingredients for successful formative e-assessment which included the immediacy of feedback effecting learner motivation.

Another example of an e-assessment tool motivating learners is badging, described by Broadfoot et al. (2013e). Badging is an alternative accreditation system arising from online communities as members validating each other's knowledge, skills or experience via the award of a visual icon. Another benefit of badging is that it can be used to help learners' visualise possible learning pathways. One specific study by Neumann (2010) found that when learners in HE evaluated a science simulation assessment, it was found that it engendered confidence and the author concluded that it held benefits for motivation.

Gee and Schaffer (2010) found that in gaming-style assessments, players were motivated because a sequence of activities gradually increased in difficulty so that players were constantly working at the edge of their abilities.

3.9.2 Learning gains

E-scape (Kimbell et al., 2009) reported on an advantage that came with more reliance on process driven activities for learners being more acquisition of soft skills with a verifiable collection of evidence generated by this type of activity. Broadfoot et al. (2013b) felt that questions answered by students using mobile devices or EVS (electronic voting system / clickers) promoted real-time feedback, collaborative interaction and reflection. While JISC (2007) discussed how e-assessment may illuminate skills of critical thinking, effective decision making, collaborative skills and practical problem solving.

3.9.3 Learner Confidence

Beevers et al. (2011) identified a further beneficial effect of e-assessment on the learner as being how feedback from a computer is non-judgemental, so the learner can explore knowledge and skills privately and comfortably.

3.10 Quality

This section contains descriptions of the issues around the quality of e-assessments. It includes a discussion of the value added by e-assessment, as compared to paper assessments, and of issues around reliability and validity.

3.10.1 Added value paper to e-assessment

Three papers were found that looked at what added value e-assessment gave over paper tests. Wylie and Dolan (2013) described how technology could provide quality data which could be categorised in terms of student learning, misconceptions and cognition to improve the information that learners receive. Furthermore, Boyle (2006) stated how e-assessment had the potential benefit over paper assessments of providing feedback tailored to the learner. Neumann (2010) felt that the impetus for using e-assessment was large class sizes and that e-assessment allowed tutors to track and record student activity.

Pachler et al. (2009) also considered what 'e' added to formative assessment, and his findings incorporated what Kennewell (2001) described as the 'affordances' of the technology. Pachler et al. found five main advantages of formative e-assessment over formative assessment: speed, storage capacity, processing, communication, and construction and representation:

Speed

- Speed of response is often important in enabling feedback to have an effect

- The ability to give feedback quickly means that the student's next problem solving iteration can begin more quickly.

Storage capacity

- The ability to access very large amounts of data (so appropriate feedback/additional work/illustrations can be identified).

Processing

- Automation - in some situations the e-assessment system can analyse responses automatically and provide appropriate feedback

- Scalability – can often be the result of some level of automation

- Adaptivity – systems can adapt to learners' needs / skills.

Communication

- Often the advantage of the 'e' is that it enables rapid communication of ideas across a range of audiences, and the technology allows this range to be controlled, it can be just one person, a group, a class or more

- Aspects of communication can be captured and given a degree of semi-permanence

- This semi-permanence supports the sharing of intellectual objects.

Construction and representation

- Representation – the ability to represent ideas in a variety of ways and to move and translate between these representations. E.g. for learners who are not highly literate, the visual nature of the screen can increase motivation (Richardson et al. 2002)

- Technology can support learners in the representation of their own ideas

- Through representation, technology enables concepts to be shaped and this helps learners develop their meaning

- In representing their ideas in digital artefacts learners open up a window on their thinking.

- Mutability – shared objects are not fixed, they can change or be changed with ease.

Broadfoot et al. (2013e) outlined their perceived benefits of using digital technologies for e-assessment: they can provide opportunities for submitting evidence via a range of media; offer more personalised assessments (including prediction modelling); support the integration of summative assessments into learning activities in order to support learner reflection and development; and provide online simulations and environments that are more authentic and relevant.

3.10.2 Validity

Dennick (2009) discussed how assessment principles, such as reliability and validity, were just as important in e-assessment. He described how in fact e-assessment offered potential for new types of questions and formats which could be used to enhance reliability, validity and utility.

The short history of computer-based assessment (CRLI, 2009) shows that technology based testing does not always equate to results obtained for traditional paper tests. There are three main areas that are seen to be potential contributors to these differences.

1. solving problems displayed on the screen requires different cognitive processes from those required when working on paper

2. task types usually associated with paper tests may not always transfer to an electronic medium and

3. the ability of students to demonstrate ability in the assessed skills may be influenced by, or restricted to, their level of IT application skills.

Others that have commented on these issues around validity include Liu et al. (2001) who described evidence for increased validity as follows:

- assessments were more closely matched to the material being taught

- presentation of more than one medium of information seemed to aid the students' recall

- questions reflected real-world situations more accurately and

- students seemed to learn more in their assessments which helped them as they continued their studies.

Swithenby (2006) warned that 'learning experiences that are increasingly mediated through screen activities should be assessed using similar media.' In a study of on-screen mathematics assessments, Hughes et al. (2011) reported how, in interviews, learners talked about their preference for working on paper over the computer, and how this was 'familiar' and 'more natural'. In order to make e-assessments valid, there is a need to make sure that any onscreen formats are familiar to learners and to be aware of how learners show their working, for example, using an onscreen notepad like Overrite.

Dennick (2009) found that when using adaptive testing, questions varied from individual to individual, but if the range of these variables was within agreed boundaries, the reliability of the test should not be greatly compromised'. Dennick also commented on two specific types of validity:

– *Face validity*: Does it seem like a fair test to candidates? This is important with e-assessment as learners may be unfamiliar with its processes.

– *Content validity*: Can it be enhanced by using animations, video and sound, hotspot questions, dragging labels over pictures and simulations?

However, Threlfall (2007) argued that construct validity was most at risk when considering e-assessment and that the key considerations were how the assessment and the teaching were related and how the mode of assessment enabled the learner to demonstrate their understanding. For e-assessment to be valid we must accept that cognitive processes used when working onscreen and on paper may not be the same. Developers should not try and replicate paper testing, but each mode of assessment should exploit the affordances of that mode, be it paper or screen, to create an authentic experience for the learner. So, for example, the creation of an onscreen ruler which can be dragged over a line to 'measure' it is not a valid task to ask a student to do onscreen (that is replication of paper assessment for all the wrong reasons i.e. administrative not pedagogical); whereas using technology to create and manipulate a graph may well better represent the type of thinking that the student developed and used in the learning of that construct, resulting in a more valid task and valid interpretations of the assessment outcomes.

3.10.3 Reliability

Some affordances or benefits of e-assessment provide opportunities for increased reliability of assessment, particularly in relation to the marking of objective questions which can be automatically marked (Meadows and Billington, 2005). The possibility of human error is also removed when totalling scores.

3.11   Implications for the implementation of e-assessment
In this section, issues raised in the literature relating to what should be considered when putting e-assessment in place are outlined.

BECTA (2003) identified a number of obstacles to be overcome in order to improve the use of ICT in classrooms:

• lack of access to appropriate equipment

- lack of time for training

- lack of models of good practice

- negative attitudes

- computer anxiety

- fear of change

- unreliable equipment and lack of technical support.

The following sections relate to:

a) The source of the impetus for using e-assessment

b) Resources and technical support

c) Teacher training and CPD (including teacher orientation to technology and their ped-agogical approach) and

d) Time and frequency of formative e-assessment.

3.11.1 The source of the impetus for e-assessment

Boyle et al. (2011) analysed three large scale UK e-assessment initiatives and concluded with advice for doing better e-assessments including these that are relevant to the ASSIST-ME project:

1. Where e-assessment is part of a policy initiative, make it a central part of the initia-tive, rather than an after-thought or peripheral concern.

2. Organisations should have a definite, positive reason for doing e-assessment; not just a vague sense that it may address the weaknesses of traditional, pencil and pa-per approaches. However, organisations should also be realistic and not go over-board with e-enthusiasm.

3. Organisations should find out their users' orientations to e-assessment; some may be conservative (such as schools) whereas others (for instance, employers) might see e- assessment as an essential tool to help them to implement education and training.

Feldman and Capobianco's (2008) literature review found that teachers choose to use tech-nology (or not) in the classroom based on their own beliefs and confidence in using the technology themselves. Kimbell et al.'s (2009) evaluation of E-scape e-portfolio system similarly found that success was often related to the enthusiasm and skills of individual teachers and that barriers included resistance to change, political complications of introduc-ing new systems into established organisations and high teacher workloads.

So there is some evidence that the impetus for e-assessment take-up comes from school – level or class-level. Unfortunately this conflicts with advice coming from more than one source which recommends that top-down implementation of change, with senior manage-

ment support and financial commitment is more likely to result in successful implementation, especially where large scale use of e-assessment is required (JISC, 2007).

3.11.2 Resources and technical support

The implementation of e-assessment brings with it some specific issues that go beyond those relating to e-learning.

Much of the literature investigated originated in Higher Education (HE) or Further Education (FE). Resources and technical support in schools may be different to those in HE, so this should be kept in mind when interpreting these findings; resources and technical support are required for the implementation of e-learning and e-assessment.

Concerns were raised in the literature about the possible shortfalls in terms of resources and technical support required to enable the implementation of e-learning. An evaluation of the e-portfolio used in schools, E-scape, (Kimbell et al., 2009) identified barriers and enablers common to e-assessment/e-learning systems; enabling factors included minimal network disruption. The impact on the capacity and speed of the school network was a priority issue for centres, as was the non-uniform hard and software provision in schools across the UK (a problem which could be exacerbated when we consider Europe as a whole).

JISC (2007) described a need for technical support and resources including a programme of technical and pedagogic support for teaching staff, interoperability with other systems in institution and shared item banks. There are also potential problems to consider when using summative e-assessment including loss of data, verifying candidate's identity and training for e-invigilators.

Also specific to e-assessment is the need for authoring tools and support for e-assessment developers, especially if they are making the transition from paper assessments to the use of technology. Quellmalz (2009) described the use of tools to guide the process of item writing and item banks that enabled efficient development and assembly of items.

Dennick (2009) provided practical advice on how to implement e-assessment for a course from which we can learn lessons including identifying clear roles for all staff when scaling up e-assessment and considering the financial demands of implementing e-assessment.

3.11.3 Continual Professional Development (CPD) and teacher training

A number of themes arose in the literature relating to Continual Professional Development (CPD) that may be required for the implementation of the ASSIST-ME e-assessments:

- The use of technology
- The use of e-assessment
- The processes of formative assessment
- The links between formative and summative assessment
- The pedagogic approach of IBE and associated competences

- And, importantly, the links between these characteristics for each of the STM subjects.

Assessing the need for CPD would be a complex task, as participants bring with them a wide variety of experience of each of these areas, as well as a variety of orientations and attitudes towards them. It was argued that e-assessment can enhance a learner's experience if assessment is closely aligned to the pedagogic approach used (JISC, 2007) suggesting that an understanding of teachers' orientation to technology, to formative assessment and IBE practices is necessary to support the use of an e-assessment which values these three aspects. Staff need support during the transitional phase to manage traditional and new methods simultaneously (JISC, 2007).

Wylie and Dolan (2013) reported that the use of items which reported misconceptions to the teacher and acted as a prompt for student/teacher discussion did not require significant change to practice; but they warned that teacher readiness and teacher development must be considered. Work on the E-scape e-portfolio project (Kimbell et al., 2009) reported that it was more effective to support and extend teachers' existing skills than impose radical change.

There was much advice to be found in the literature, the summary of the messages being that the availability of teacher professional development and the release from workload in order to take-up CPD is critical to success (e.g. Broadfoot et al. 2013b, Whitelock, 2006).

3.11.3.1 Use of technology

It was reported that teachers needed to feel that the use of computers in the class was manageable (ALTA, 2009). This suggests that personal orientation, preferences and having the skills and confidence to take ownership of and use technology varies considerably across teachers.

3.11.3.2 Use of e-assessment

Good CPD is needed to enable teachers to take advantage of the rich reporting capabilities of technology (Beevers et al., 2011). If teachers understand the affordances of the technology, then they have the power to make decisions about when it is valid to use technology and when to use traditional methods.

3.11.3.3 Formative assessment

The message came from Feldman and Capobianco (2008) that for teachers to really take on formative assessment and for it to make an impact, they need:

- time and opportunities to engage with the software and hardware

- to understand the items and their relationship with learning and pedagogic methods and

- the opportunities to collaborate with other interested teachers.

A core component around which there is much difference between e-assessment and paper-based assessment is the role of the teacher and to what extent their role in formative assessment includes adaptation of pedagogy (Pachler et al., 2009). E-scape (Kimbell et al.,

2009) used existing hardware and software so that it was familiar and easy to follow. Teacher training was more focused on the collection of data and the analysis of it for formative purposes rather than on the use of the hardware and software.

3.11.3.4 IBE

Kennewell (2008) reported that, when teachers first adopted technology as part of their practice, there was a tendency for interactivity to be superficial and authoritative and only when technology was embedded in teachers' pedagogical knowledge did the technology contribute to deeper, more dialogic interaction amongst students. This suggests that teachers need to have an understanding of the technology and IBE and the relationship between them and what technology can and cannot bring to IBE.

3.11.4 Time and Frequency of formative e-assessment

Cheung (2011) reported that assessment programs that were used for more than 30 minutes a week had a bigger effect than those that were used for less than 30 minutes a week. This reflects Swithenby's (2006) concern that key issues for the success of assessment are 'the regularity and quality of student engagement, the timelines and quality of feedback and the student engagement with their feedback'.

3.12   Implications for evaluation of assessments

Halperna et al. (2012) reported on two evaluation studies of Operation ARIES!. These studies used experimental trials in which learning gains, as measured by performance on short answer questions, was compared across different types of e-learning and types and immediacy of feedback were investigated. The use of short-answer questions to measure performance gains could be criticised for a limited view of what benefits the learning types and feedback types could bring, for example, these performance gains seemed not to include IBE or skills.

Evaluations of the ALTA system (2009) on the other hand, considered wider issues including practical concerns like:

- the ease of use

- teacher training

- manageability for pupils and teachers,

- the ability to use computers in class

as well as educational issues including

- engagement from pupils,

- whether formative assessment was promoted

- how teaching and learning was supported

- whether mathematical skills were developed

- if there was any impact on pupils' ability and

- if self-assessment was enabled.

3.13   Exemplars

In this section examples of e-assessments are listed and described. Criteria for selection of which examples to include as exemplary are:

- e-assessments which are or have been in use and are well-established

- e-assessments which have an element of success i.e. provide some evidence of good practice.

- e-assessments which meet at least one of the ASSIST-ME aims for assessment:

    o   support formative assessment

    o   enable both formative and summative assessment

    o   cover STM subjects

    o   focus on inquiry-based education

    o   focus on competences related to inquiry-based education or

    o   are relevant to primary and/or secondary education.

JISC (2007) made suggestions about the experiences and needs learners could acquire at different stages of learning; this gives an indication of the types of experiences that may appear in the exemplars from different levels of education:

5-11 years	11-14 years	14-16 years
web-based interactive multimedia learning resources and games, drill and skill quizzes, e-profiling of early years development, high quality, web-based AfL, SATs based on online assessments and exemplars	BBC mobile bite-size quizzes for learning on the move, gaming, online banks of SATS, multimedia resources for learning, innovative games-based assessments	online assessment materials for gifted and talented learners, virtual world simulations testing skills in context (ICT and science), opportunities to personalise their learning using on demand online testing

Exemplars are presented by subject, or combination of subjects.

3.13.1   Mathematics

3.13.1.1 HE statistics simulation

Neumann (2010) described simulations used for summative assessment of HE statistics. The use of technology for these simulations closely reflected how technology was used in the teaching and learning of statistics. The use of simulations enabled students to have control over their actions and make decisions which had impact.

3.13.1.2 NRich Mathematics

This team at the University of Cambridge works to enrich the experience of the mathematics curriculum for all learners aged 5 to 18 by offering online enrichment materials (problems, articles and games) to be integrated with every day practice. The main focus is on the development of mathematical thinking and problem-solving skills and many of the activities show rich mathematics in meaningful contexts.

Activities are accessed via the NRich website and some of them exploit the affordances of the technology, including digital learning objects such as simulations and interactive tasks. The resources are organised such that some activities are categorised by level of difficulty: warm up, try this next, think higher, explore further. This gives the learners support in choosing which way to proceed.

The interactive, animated tasks on the NRich website www.nich.ths.org could be used for formative or self-assessment. Examples of interactive tasks include animations for tasks using peg boards, Cuisenaire rods, cogs, shapes, grids and Carroll diagrams; such tasks enable learners to attempt tasks again and give instant feedback. They could be used as stimuli for IBE, particularly where stimulations and interactive tasks are used.

3.13.1.3 Misconceptions in mathematics

Wylie and Dolan (2013) reported on the creation of a bank of items for high school mathematics and science teachers that drew on the misconception literature (Wylie and Ciofalo, 2008). Each multiple-choice item that was developed drew on at least one previously identified student misconception, so the formative e-assessments were through MCQs, with some distractors relating to particular known misconceptions. The advantage over paper formative assessment was that the process of translating student responses into diagnoses of misconceptions was automated. These question types have different implications for item development than those which are not misconception based (Wylie and Dolan, 2013).

3.13.1.4 Progression assessment

Like the misconceptions-based assessments described above, Arieli-Attali et al. (2012) identified an assessment which aimed to provide quality data to support decision making to inform the next instructional steps and improve the information that learners received. This middle school mathematics project focussed on the progression between levels of understanding, rather than just on categorising the learner according to at which level they sat (CCSSO, 2008). One interesting aspect of this tool is how it uses two kinds of assessments which could also feed into the gathering of information for formative and summative assessment purposes. These two kinds of assessments are: locator assessment (computer delivered and placing the learner within three learning progressions) and incremental tasks (which explicitly targeted a transition between levels, rather than the levels themselves).

3.13.1.5 The ASSISTment system

Pellegrino and Quellmalz (2010) looked at the ASSISTment technology-enabled assessments which use a pseudo-tutor for middle school level mathematics.

The system used scaffolding questions, optional hints, and buggy messages (specific feedback given after student errors) for each item. Students are eventually guided to reach the correct answer; scaffolds and hints are limited to avoid giving away the answers. Teachers

receive feedback on student and class progress both on general summative measures (for example, percentage correct), on more specific knowledge components, and on formative aspects. This e-information available to teachers not only allows them to analyse individual and group performance, but the enhanced information, afforded by the technology, can feed into adapting teachers' pedagogy through the use of formative information.

3.13.1.6 Abacus Evolve

Pearson's (2013) Abacus Evolve mathematics programme provides online mathematics games. When originally introduced the materials included an interactive CD ROM, a Talk Maths CD ROM for pairs of children to use, a Solve The Problem CD ROM for pairs or groups and individual practice software. The materials designed for pairs and groups of children to work on together may compliment an IB approach and may also yield information when assessing competences.

3.13.1.7 Centre for Mathematics, Science and Computer Education

The Centre for Mathematics, Science and Computer Education (whose purpose it is to improve mathematics, science, and computer education programs in the States) offers links to many interactive mathematics tools, including ideas for teaching and assessing mathematics with examples of online support for pupils. While these are not comprehensive systems like many of the other exemplars that we have described, the electronic mathematics resources do lend themselves to use in the IB classroom. Four specific tools are described below.

1. The National Library of virtual manipulatives (http://nlvm.usu.edu/en/nav/category_g_2_t_1.html_e.g.) Here digital learning objects such as abacus, fractions, number lines, bar charts and Venn diagrams can be found.

2. Online mathematics manipulatives http://www.ct4me.net/math_manipulatives.htm . Here learners can submit answers and, if they are wrong, they get instruction and examples of worked through answers. It covers a large range of mathematical content areas for students aged 5 to 16.

3. Visual mathematics learning http://www.visualmathlearning.com/. This has onscreen exercises for practice in mathematics, some of which have useful visual clues to support learners with answering questions.

4. Math cats http://www.mathcats.com/. This site also has interactive mathematics activities, some of which link to teaching in a more formative way and some of which link to more summative-style assessment. One feature allows learners to create graphs, http://nces.ed.gov/nceskids/Graphing/ including bar charts, pie charts and line graphs.

3.13.1.8 Cognitive Tutor

Described as 'adaptive curricula', Cognitive Tutor software, http://www.carnegielearning.com/specs/cognitive-tutor-overview/ , was developed around an artificial intelligence model that identifies weaknesses in each individual student's mastery of mcompetencesathematical concepts. It then customises prompts to focus on areas where the learner is struggling and sends the learner to new problems that address those specific concepts.

Online activities include:

- multiple representations (these can be expressed numerically or display problems graphically)

- worksheet prompts to convert problems into mathematical expressions

- interactive examples (with step by step instructions for learners)

- flexible sequencing (for teachers or administrators to determine)

- pre and post-tests (a pre-test can be diagnostic and set the pace for further instruction)

- immediate feedback, including giving learners the opportunity for self-correction; the programme recognises the most common errors and misconceptions and responds appropriately and

- a 'skillometer' which indicates the journey to mastery for learners and teachers.

The flexible sequencing and pre and post-tests lend themselves to formative and summative assessment practices and, with the personalisation of the content, the system also allows for self-assessment.

3.13.1.9 ALTA

The ALTA (Adaptive Learning, Teaching and Assessment) system targets KS1 – 3 mathematics (age 5 to 16 years). It was designed to support and promote formative assessment, to inform self-assessment and to inform teaching through the collection of longitudinal records of pupil performance. It includes built-in information resources e.g. question banks mapped onto curricula, and all assessments are adaptive. Teachers can see pupil and class profiles; diagnostic analyses can show trends. The curriculum is represented on a 'STLC grid' which shows the subject, topic, level, criterion (unique learning objective) enabling questions across a range of difficulties. Developed and used in Northern Ireland, ALTA has been independently evaluated fifteen times over five years including five CCEA (Northern Ireland Curriculum Authority) evaluations which all report positively. This approach could be adapted for use with competence-based curriculum and the use of longitudinal data could feed into both formative and summative assessments, with the trends from the diagnostic analysis feeding into the teachers' adaptations of pedagogy.

3.13.1.10   Bioware and Atari computer game

In the Bioware and Atari computer game, 'Neverwinter Nights', players have to improve their literacy and numeracy skills in order to progress. Completed tasks are banked in an e-portfolio for assessment which was shown to significantly improve success rates in basic and key skills assessments (JISC, 2007). This is an illustration of how gaming-style e-assessment can motivate learners and a series of formative assessments can feed into a summative result.

3.13.2 Science

3.13.2.1 SimScientists

SimScientists is a set of simulation-based science assessments used in middle school science classrooms. The system uses simulations to prompt curriculum-embedded formative assessment. The system identifies types of errors and follows up with increasing levels of feedback and coaching for learners, from identifying that an error has occurred and asking a student to try again, through explaining the concept, to demonstrating and explaining the correct answer (Quellmalz et al., 2012). This model supports IB learning.

SimScientists can also be used as a summative, benchmark assessment providing evidence of middle school students' understanding of ecosystems and inquiry practices (having completed a regular curriculum unit on ecosystems). It illustrates ways that assessment tasks can take advantage of simulations to represent generalisable, progressively complex models of science systems. It present significant, challenging inquiry tasks and provides individualised feedback and customised scaffolding. It claims to promote self-assessment and metacognitive skills which could be in line with IB-related science competences.

Quellmalz et al. (2012) described how SimScientists linked the targets to be assessed with evidence of proficiency on them, and with tasks and items eliciting that evidence (Messick, 1994; Mislevy and Haertel, 2007 in Quellmalz et al. 2012). The process begins by specifying a student model of the knowledge and skills to be assessed. The SimScientists assessments used the evidence-centered design method to align the science content and inquiry to be assessed, to scoring and reporting methods, and to the specification of the assessment tasks and items.

3.13.2.2 Operation ARIES!

Operation ARIES! and later named Operation ARA! (Koenig, 2011) was designed to assess and teach critical thinking about science which relates to specific science competencies. It uses intelligent tutoring and makes use of avatars. Students watch videos and receive communications through email and text message. The approaches involved in the modules are related to IB teaching and learning. It includes three types of module:

1. Interactive training: students read an e-book and after each chapter they are quizzed with multiple choice-type questions. Students receive feedback and tutorage from two avatars from the program throughout.

2. Case studies: students are expected to apply what they have learnt in the previous modules.

3. Interrogation: learners are presented with inaccurate science information through the medium of newspaper headlines and television news channels and must ask questions to ascertain the truth.

Through all three modules, key principals of learning are included, such as:

- Self-explanation- the learner communicates the material to another automated student

- Immediate feedback- through the tutoring system

- Multimedia effects- aiming to engage the student

- Active learning- students engage in solving a problem

- Dialog interactivity- students learn by engaging in conversations and tutor groups and

- Real life examples- intended to help students transfer what they have learnt in one context to another.

3.13.2.3 Zydeco Mobile Application

Zydeco(Delen et al., 2012) is a mobile application to support learners in creating scientific explanations using a claim-evidence-reasoning model. The tool supports collection of multi-media data which can be in the form of audio notes, videos and photos from both inside and outside of the classroom. Learners can use evidence coming from other learners as evidence to support their claims. Evaluations claimed that it supported pupils in the 6th grade to develop scientific explanations including creating hypotheses and linking data to them.

3.13.2.4 PISA and NAEP

PISA and NAEP are two examples of where e-assessment and its affordances are being introduced into formerly paper-based assessments. Pellegrino and Quellmalz (2010) reported on how the 2006 PISA pilot tested a computer-based assessment of science to test knowledge and inquiry processes not assessed in the paper-based booklets. The 2009 NAEP Science Framework and specifications drew upon ETS science simulations work (CCSSO, 2008) and other research to develop their rationale for the design and pilot testing of interactive computer tasks to test the students' ability to engage in inquiry practices. These innovative items were included in the 2009 NAEP science administration (Pellegrino and Quellmalz, 2010).

3.13.3 Mathematics and Science

3.13.3.1 Hungarian diagnostic assessment

An example of how e-assessment is being developed for national use can be found with the Center for Research on Learning and Instruction (CRLI, 2009) in Hungary. Here a networked platform for diagnostic assessment is being developed from an online assessment system. The goal is to lay the foundation for a nationwide diagnostic assessment system for grades 1 through 6. The project will develop an item bank in nine dimensions (reading, mathematics and science in three domains).

3.13.3.2 Scholar

Scholar (www.scholar.hw.ac.uk) is one of the largest online learning programmes in the world with over 80,000 registered students in Scottish secondary schools. It provides online educational resources and a 'virtual college' support network. There are separate courses for biology, physics and chemistry and for mathematics. The science programmes contain some animated graphs with associated questions (some referred to as investigations) and the mathematics programme has step by step demonstrations, for example, on how to construct a pie chart.

It provides opportunities for independent learning plus formative assessment, with pupils working on repeated practice with immediate feedback. Item types include multiple choice questions, short answer and extended answer questions alongside a variety of materials including:

- sets of learning points for revision

- e-learning content for each topic area (which could be in the form of static text, diagrams and/or graphs) and

- revision planners for pupils.

Results of the end of topic tests go to teachers. Staff training is tailored to the school's needs.

3.13.3.3 SAM Learning

SAM learning (www.samlearning.com) is an online revision and test practice package for mathematics and science designed to mainly support individual revision which is often completed at home. The website claims that 10 hours on SAM Learning improves student achievement by 1 GCSE grade. Motivation for students comes from choosing avatars, 'playing' against their friends and seeing their friend's progress (with the aim of motivating them to do more revision). Activities viewed on the demonstration programme included onscreen versions of paper items (some with drop down menus), some drag and drop items and some fill in the blanks items. The interesting aspect of this is that it is a successful self-assessment package where no teacher input seems necessary.

3.13.4   Technology

3.13.4.1 National Curriculum Test in ICT for 14-year olds in the UK

Boyle et al. (2011) described the development and pilot of the KS3 ICT test. The test was radical in that it contained several novel features and required that learners solve problems in a virtual world. It assessed ICT capability although the initiative was criticised because the construct was not widely understood nor explicitly aligned to existing widely understood constructs of ICT competence.   Finally the test was downgraded from a high stakes summative statutory to a formative test and redesigned as free-standing assessment tasks (QCDA, 2011).

3.13.4.2 NAEP technology and engineering

The new 2014 Technology and Engineering Literacy Framework for NAEP will be entirely computer administered and will include specifications for interactive, simulation-based tasks involving problem solving, communication, and collaboration related to technology and society, design and systems, and information communications technology (Pellegrino and Quellmalz, 2010).

3.13.4.3 E-scape

Kimbell et al. (2009) evaluated the use of e-portfolios for design and technology, geography and science. E-scape was designed to capture the process that learners go through, including capturing collaboration. It had the facility to capture a timeline of learner activities. Learners could record their own ideas alongside any justifications for their ideas or actions. The

intelligent tutoring included teaching elements and the timeline could highlight points in the learner's development e.g. when designing a guitar, learners looked at videos and articles, it allowed students to comment on these articles or stimuli and the timeline captured how this affected the development of their design. The digital form of the e-portfolio meant that there were more options for students to upload material e.g. the use of film, photos, maps and web links.

3.13.4.4 BTEC in IT skills

JISC (2007) discussed the BTEC intermediate and advanced award in IT skills for learners in small businesses from Edexcel. The course is delivered via the internet with tutor support online. The skills-based course allows learners to progress at their own pace and has no formal examinations. On-going assessment is assimilated into the structure and content of the course. Learners complete tasks as evidence of achievement and self-assessment exer- cises at the end of each unit allow them to obtain formative feedback.

4 Conclusions and implications

The ASSIST-ME proposal states that 'ASSIST-ME will develop formative assessment methods that (1) fit into everyday classroom practice, (2) provide qualitatively oriented descriptions and monitoring of competence-oriented, inquiry-based learning processes, and (3) can be combined with existing summative assessment requirements and methods used in different educational systems. The assessment methods will be developed to capture both general competences and disciplinary process competences such as science investigations and authentic problem solving.

The development and design of these methods will be based on existing research on formative and summative assessment, on current research-based understandings of competences in STM, and on previous and on-going EU projects on inquiry-based education (IBE).'

Following some reflections on the application of the methodology and the analysis of the data, the conclusions are presented with reference to the objectives set out in section 1.3.

1. Identify existing relevant digital assessments

2. Through the literature, identify theories and models which are relevant to the development of such digital assessments

3. Identify strategies used in the evaluation of the models which could inform good practice

4. Identify implications for the development of the digital assessments relevant to the aims of ASSIST-ME.

4.1 Reflections on the methodology and the process of searching

The literature search involved a quite general review of literature on e-assessment rather than looking specifically at the narrow focus of e + IBE + STM + competences, as this field yielded very little information. However, these broader findings do relate to the specific ASSIST-ME focus.

The literature search revealed that the majority of reported use of e-assessment found was in the Higher Education (HE) and Further Education (FE) sector. While the focus of the ASSIST-ME project is for the primary and secondary phase, findings showed how it was possible to use HE as a proxy for the primary and secondary sectors and that general principles of e-assessment used in HE/FE could be used in this project.

Dermo (2009, in Voelkel, S. 2013) reported on how 'e-assessment is widely accepted by students as part of their university studies and they generally feel that it had a positive impact on their learning'. We are reminded that HE is used to e-assessment and e-learning, Dermo here saying this is in part due to technology having been embraced.

A small amount of the findings was specifically related to IBE. However, through reading, it became clear that in order to assess IBE (and related competences), there's a need for learners to engage in IBE as part of their assessment and so ideas and e-learning and as-

sessment together could feed into this process. There do exist conceptions that e-assessment is more compatible with objective, multiple-choice type questions, but actually our findings show that this conception is not correct and exemplars have shown how e-assessment and IBL are compatible.

Originally the literature search was focused on the last few years, but it became evident throughout the research that changes in e-assessment practices were slow and minor and so it was in fact appropriate to look back over the last 15 years.

4.2 Conclusions specifically related to objectives

4.2.1 Objective 1: Theories and Models relevant to development of digital assessments

A variety of theories and models for e-assessment have been identified. This section aims to draw out the advantages of e-assessment, as compared with paper/traditional tests, from these themes to establish good and interesting practice. These themes are followed up under objective 4 (section 4.2.4), where implications for the development of e-assessment for ASSIST-ME are discussed.

4.2.1.1 Teaching/Learning/Assessment Link

Learners benefit when teaching, learning and assessment are linked and various sources in this review have demonstrated how e-assessment can facilitate this. Models identified in the search have blended teaching and short, repeated, formative tasks and have demonstrated how a sequence of these formative tasks can combine to give a more summative assessment. Gaming-style assessments can be a model where the teaching is all on-screen and through a series of smaller tasks (the formative element), learners, working at the edge of their ability can move towards mastery of skills (the summative element).

4.2.1.2 Stimulus types

One of the affordances of the technology is the types of stimulus that can be provided for learners to work with during e-learning and e-assessment. Digital learning objects can be used as a stimulus and learners can manipulate them to give them some grounding for their next actions. Worked exemplars can be included which support learners in knowing how to proceed with a problem/investigation. A variety of models for these has been found from examples where the entire process is revealed, to ones where hints and clues can be given to students or where only the next small step is revealed. A broader range of real-life scenarios and simulations are often more easily replicated on-screen. These cannot only offer more valid assessment stimuli, but can also increase students' motivation.

4.2.1.3 Feedback

Many successful models of e-assessment include elements of feedback, from the simple to the complex. The immediacy of the feedback offered by e-assessment has been shown to be a motivating factor for students both in terms of continuing with their learning/revision/understanding and increasing the likelihood that students undertake e-assessments out of choice. E-assessment feedback comes in many forms and can be a form of learning in itself with feedback models, for example, linked to intelligent tutoring offering links to further learning.

4.2.1.4 Adaptivity

E-assessment enables a more interactive approach for learners. It seems that adaptive models not only personalise the learning and assessment, but act in a supportive way so that learners are more motivated and their time and effort is more focused on their ability and level of understanding and skills. These aspects of feedback and interactivity lend themselves to group work and peer assessment, both of which can be features of IBE.

These interactive and adaptive features that can come along with e-assessment can lead to more learner autonomy. It can also lead to improved self-assessment as learners are more likely to take notice due to the immediacy of the feedback offered through e-assessment. The on-demand nature of some formative and summative e-assessments has been shown to increase student motivation also.

4.2.1.5 Data for formative and summative purposes

Large amounts of data can be collected with e-assessment. This may be just scores, as with a paper tests, but it may also give group scores, compare scores to previous cohorts, compare sub-groups of the whole cohort and give information about progress. As well as this type of data, e-assessment can also give insights into how learners approached the task. This mass of data can enhance the teachers' understanding of the learning and act as a strong contributor to formative assessment and adapt the behaviour of the teacher and the learner. This is particularly relevant with this project, as it can capture individual student's input and can collect evidence of how learners approach problem solving including the sequences they used, the strategies they used, the number of attempts a learner took and the amount of time taken on different sections of the activity.

The use of learning analytics can feed into predictive modelling, user profiling and adaptive learning and so support a formative approach.

4.2.1.6 Re-submission

One aspect that e-assessment allows for, and which can feed into meaningful, formative assessment, is the ability to re-submit answers. Original inputs can still be captured for use by the teachers so that they are aware of how easily a learner may have arrived at an answer, but for learners, the immediacy of this feedback and the opportunity to resubmit can make the process and learning more relevant than waiting a period of time to see the results. The immediacy of feedback with e-assessment allows learners to move directly on to the next stage which has advantages in formative assessment and with self, peer and diagnostic assessment. Learners can also submit a confidence rating with their answer which adds further meaning to the level of information gathered.

4.2.2 Objective 2: Strategies used in the evaluation of the models which could inform good practice

In the literature that was reviewed, there were few examples of how e-assessments had been evaluated. A model for development of e-assessments did not seem to follow the usual pattern of design – trial – evaluate and so few cases of evaluation techniques were identified.

In the evaluation of Operation ARIES! (Halperna et al., 2012) learning gains were measured in terms of short answer responses. This was compared across:

- different types of e-learning and

- immediacy of feedback.

In the evaluation of the ALTA system (Adaptive Learning Teaching Assessment for mathematics KS1 – 3) a less limited view of learner gains was adopted. It analysed:

- practical concerns: ease of use, teacher training, manageability for pupils and teachers and ability to use on computers in class and

- educational issues: engagement from pupils, whether formative assessment was promoted, how teaching and learning was supported, whether mathematical skills were developed, if there was any impact on pupils' ability and if self-assessment was enabled.

4.2.3   Objective 3: Existing relevant digital assessments

This report has identified both individual elements of digital assessments and complete assessment programmes which may comprise a number of elements including learning. The complete programmes considered in the review are ones that are established; some are used organisation-wide and some are national programmes. The exemplars are also in the subjects considered in this project: science, mathematics and technology.

Finding existing programmes has shown how e-assessment can be integrated into the learning process and be successful in terms of motivating learners, assessing competences that are more difficult to assess using paper tests and can include elements difficult to replicate in a paper test, for example, simulations, videos and scenarios that lend themselves more readily to IBE and competence-based skills.

4.2.3.1 Mathematics

The mathematics exemplars included the use of questions, simulations, enrichment activities, activities designed to assess misconceptions and online games. They were designed to feed into summative assessment, formative assessment, self-assessment and diagnostic assessment. Some were adaptive in nature. They were aimed at a range of age groups: 5-16 years, 5-19 years, middle school, high school, and Higher Education.

Some reflect how technology is used in teaching and learning, others use digital learning objects or online/virtual manipulatives. Some use responses for diagnostic purposes, identifying where a learner is struggling and directing them to new questions to address specific concepts. Others focus on learner progression to inform the next instructional steps. Some involve a pseudo-tutor with hints and feedback after errors; with these detailed feedback is also passed to teachers. Certain exemplars bank completed tasks in e-portfolios and others inform teaching through the collection of longitudinal records.

4.2.3.2 Science

The science exemplars often involve inquiry tasks and processes and critical thinking (closely linked to the ASSIST-ME focus). The exemplars found focused on summative and formative assessments. They were aimed at ages 11-19 years and middle school students.

They incorporate the use of feedback and some are interactive in nature. They utilise e-coaching as well as intelligent tutoring. Some include the facility to re-submit answers. Some employ videos and case studies and many use simulations.

4.2.3.3 Mathematics and Science

The exemplars found were mainly for use in revision and for independent learning. They were aimed at all secondary school ages. They were designed for formative assessment, diagnostic assessment and self-assessment.

One incorporated an item bank and another utilised digital learning objects and demonstrations of answers.

4.2.3.4 Technology

The technology exemplars found were used summatively as they were all high stakes qualifications and, as such, aimed at 14-19 year olds. Some aspects could also be used for formative assessment purposes and one claimed the on-going nature of the programme could be used for self-assessment.

Some demanded problem solving in a virtual world and others used interactive simulations. They utilised intelligent tutoring and online tutor support, as well as allowing students to progress at their own rate. They included the use of a timeline of student activities and an e-portfolio.

4.2.4 Objective 4: Implications for the development of the digital assessments relevant to the aims of ASSIST-ME

Information gathered through this process was used to form a set of recommendations to use when considering the development of e-assessments. The aspects of this that are relevant to developing formative and summative assessments in IBE in STM subjects are presented here. A key message with regard to making good use of e-assessment is to exploit the affordances of the technology and not to just translate a paper test to an online version.

4.2.4.1 Exploiting the affordances

The affordances of the technology should be exploited to enable the assessments to go beyond what paper can do. This could be via the inclusion of particular elements, for example, scaffolding questions, optional hints and clues, simulations and scenarios, digital learning objects, all of which can support the learners to understand which way to proceed with IBE. E-assessment has the ability to be interactive which can be an advantage when using it in IBE. It could be argued that cognitive skills can be assessed on paper whereas to assess inquiry skills paper is not a suitable format.

Interactive tasks and simulations can give intrinsic, visual feedback to allow the learner to make decisions about the accuracy of their work and when they are ready to submit an answer (if that is necessary). However, there is a need to avoid 'random button pressing'. In-

teractive tasks should not be superficial or authoritative, but should give learners control and contribute to deeper, dialogical interaction among students.

4.2.4.2 E-learning and e-assessment relationship

E-assessment should be an integral part of the pedagogy and be closely linked to learning. The increase in learner control afforded by the technology allows for closer integration of teaching, learning and assessment. E-assessment loses its impact when it is just an add-on to usual classroom practice. It has also been found that the impact of e-assessment is insignificant where little time and attention is placed on it. With this in mind, it must be stressed how important it is that the format and functionality of the e-assessments are familiar to the learner. Many e-programmes firstly develop a good learning package and then add in the assessment element.

The relationship between e-learning and e-assessment is key (as is the relationship between learning and assessment) to the success of formative e-assessment. It is a matter of validity that the two need to relate to the same constructs. To ensure that the outcomes of e-assessment are used validly, there is a need to analyse which aspects of the curriculum could and do use technology in teaching/learning and to start with these as areas where e-assessment would be a valid approach.

The relationship between the teacher and student is also central to effective formative assessment, and the risks of using off the shelf e-assessments are that this relationship is distanced. Wylie and Dolan (2013) suggested that teachers should develop their own evaluative tools as part of their instructional practice. E-assessment can provide prompts for discussion, as long as teachers are skilled and/or supported in using e-assessment outputs (Wylie and Dolan, 2013).

4.2.4.3 Link to summative and formative assessment.

Having adaptive e-assessment can better support the learner with self-assessments, as more work is levelled near to their level and so more of the feedback is relevant, more detail around the edges of their ability can be gathered and so learners can get a clearer picture of where they are, identify gaps or misconceptions in their current knowledge and where they need to go next.

E-assessment can support the bringing together of formative and summative assessment, as technology gives opportunities for the blurring of traditional lines between learning, formative assessment and summative assessment (Bennett, 2002). Is the blurring of the distinction between formative and summative assessment meaningful to students? Broadfoot et al. (2013) suggest that actually it is meaningful to have a more holistic view of assessment, when assessment is authentic.

Gaming models can fulfil the purposes of both formative assessment and summative assessment (Gee and Schaffer, 2010).

4.2.4.4 Feedback

The rich feedback that can be given to students when this is integrated with e-assessment give it a distinct advantage over other types of assessment. The immediacy of receiving feedback for the learner has been shown to be a motivator, including motivating learners to

take more tests and for further learning. Another motivating factor comes about when the e-assessment is adaptive; motivation comes from when the activity is personalised to be pitched at the correct level for the individual learner.

If only one affordance of the technology were to be exploited, it should be the speed and detail of feedback that it is possible to give directly to students. Feedback should be instant, differentiated and individualised. Formative e-assessment is most effective when feedback relates to cognitive processes, not just a score or success/fail message. Feedback should to be actionable, and students need to understand what to do with the feedback.

4.2.4.5 Interventions and intelligent tutoring

E-assessment provides opportunities for interventions which are speedy, based on evidence and good pedagogic principles. The great variety of intelligent tutoring options available allow for learner style and preferences to be accounted for, to ensure that e-learning is individualised.

4.2.4.6 Teacher Continual Professional Development (CPD)

The aims of the ASSIST-ME project assessments are wide and so CPD requirements could cover a range of areas: technology use in teaching and learning; e-assessment; formative assessment; links between summative and formative assessment; the pedagogic approaches of IBE and associated competences; how each of these apply to the STM subjects; and importantly, the interrelationships between each of these aspects. Most importantly teachers need to be released from their usual workload in order to take up CPD (Whitelock, 2006). There is a need, with the introduction of e-assessment, for there to be provision for the CPD of teachers who will be engaged with the technology. There is also a need for institutions to provide technical support so that teachers can feel confident when using the e-assessment. Technical support also ensures that teachers are aware of the features of the e-assessment and of all of the data outputs from the e-assessment so that it can be best used to enhance the teaching and learning.

4.2.4.7 Implementation

Due to the nature of this type of technology and the processes involved in the development of e-assessments, it is usual that the introduction of e-assessment would need to be a top down process, that is from the management, rather than an individual teacher bringing in a novel e-assessment technique; top-down change is needed, from policy makers, as practitioners will not be able to influence change of this nature from the bottom up. Some e-assessment projects begin with one or a handful of enthusiastic teachers and succeed as small scale projects, but for e-assessment initiatives which are scaled up, a key factor in success will be the financial commitment and support of senior managers (JISC, 2007). Developers need to have a positive reason for using e-assessment and, once committed to e-assessment, they need to make it central to assessment.

During the implementation of the E-scape project, centres were concerned with support and adequate training for e-assessment alongside expectations of high cost. Dennick (2009) referred to economic issues and demands brought by e-assessment including staff to support the system, trainers, IT support and, software. Technical expertise and resources are required for the successful implementation of e-assessment (JISC, 2007), particularly to over-

come the most common problems of interoperability with existing systems, network capacity and technical and pedagogical support for staff.

4.3 E-assessment and formative assessment

Many feel that high stakes assessment is incompatible with e-assessment innovation (Boyle et al., 2011). Beevers et al. (2011) argued that when there exists tension between using e-assessment for formative assessment and summative (high stakes) assessment, the latter will not get off the ground because of political reasons. The former is ideal for innovating with e-assessment. Beevers et al. predicted that formative assessment would be the vehicle for the e-assessment breakthrough because summative assessment is risk-adverse and subject to political factors which inhibit the use of e-assessment.

4.4 Last words

The outcomes of the WP2 project will input to WPs 4 and 5 – the design of a range of combined assessment methods. The key messages contained in this document that the authors wish to highlight are summarised below.

1. When considering the validity of e-assessments, it is important to link the construct being taught/learnt and the construct being assessed. This will require an understanding of the impact of technology on teaching / learning and the construct being assessed needs to reflect this. Technology impacts on cognitive processes and subject-related thinking and so to mitigate against negative impact, technology should be integrated with existing classroom practice.

2. IBE requires learners to interact with problems and situations; replicating this in assessment is more easily done through technology than on paper because the technology offers the affordance of interaction with authentic problem solving environments, instant feedback and adapting to the learners' responses.

3. The more elaborate the e-assessment package is the more effective it is; the benefits of the technology are not realised when a paper test is migrated to screen, but can be more fully exploited when the potential affordances of e-assessment technology are fully embraced.

4. The richness of the outputs from e-assessment allows teachers to monitor individual, group and sub-groups' attainment, to see progress over time and across interventions, all of which feed into a more complete formative education process.

5. If the developers were to take up one affordance of the technology then it should be to exploit the potential for immediate and detailed feedback to the student and the associated intelligent tutoring opportunities.

6. The use of a gaming model (incremental assessments leading to mastery and progression) is an example of how to blur the distinction between formative and summative assessment, but importantly in terms of the aims of ASSIST-ME, gaming also shows how formative and summative assessment can be brought together in a IBE environment which develops and recognises competences.

5 References

ALTA (2009) Adaptive Learning Teaching Assessment (website). Available at: www.alta-systems.co.uk [accessed 11-05-13].

Anderson, R. (2002) Reforming Science Teaching: What Research says about Inquiry. *Journal of Science Teacher Education*, 13(1): 1-12.

Assessment Reform Group (2002) Assessment for learning - 10 principles: Research-based principles to guide classroom practice. Available at: http://www.assessment-reform-group.org.uk [accessed 12-05-13].

Arieli-Attali, M., Wylie, E.C., & Bauer, M.I. (2012, April). The use of three learning progressions in supporting formative assessment in middle school mathematics. *Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada*.

Beatty, I. D., & Gerace, W. J. (2009) Technology-Enhanced Formative Assessment: A Research-Based Pedagogy for Teaching Science with Classroom Response Technology. *Journal of Science Education & Technology*, 18(2), 146-162. doi: 10.1007/s10956-008-9140-4.

BECTA (2003) What the research says about barriers to the use of ICT in teaching (online report) available at: www.becta.org.uk [accessed 14-05-23].

Beevers et al. (2011) What can e-assessment do for learning and teaching? Part 1 of a draft of current and emerging practice review by the e-Assessment Association expert panel. *International Journal of e-Assessment*, 1(2).

Black, P., and Wiliam, D. (2009) Developing the theory of formative assessment. *Journal of Educational Assessment, Evaluation and Accountability* (formerly the Journal of Personal Evaluation in Education), 21 (1). pp. 5-31.

Boyle, A (2006) Evaluation of the 2006 pilot of the key stage 3 ICT test. QCA Assessment Research team.

Boyle, A., & Hutchison, D. (2009) Sophisticated tasks in e-assessment: what are they and what are their benefits? *Assessment & Evaluation in Higher Education*, 34(3), 305-319. doi: 10.1080/02602930801956034.

Boyle, A., Sceeny, P and Sowerbutts, P. (2011) A history of three e-assessment programmes in England. *International Journal of e-Assessment*, 1(2).

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013a) Rethinking Assessment, series of discussion papers. Paper 1 Transforming education through technology enhanced assessment. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013b) Rethinking Assessment, series of discussion papers. Paper 2 Integrating FA and SA through technology enhanced assessment. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013c) Rethinking Assessment, series of discussion papers. Paper 3 Exploiting the collaborative potential of technology enhanced assessment in higher education. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013d) Rethinking Assessment, series of discussion papers. Paper 4: Learning analytics and technology

enhanced assessment. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013e) Rethinking Assessment, series of discussion papers. Paper 5 Ethical issues in technology enhanced assessment. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Broadfoot, P., Timmis, S., Payton, S., Oldfield, A. and Sutherland, R. (2013f) Rethinking Assessment, series of discussion papers. Paper 6 National standards and technology enhanced assessment. Graduate School of Education, University of Bristol www.bristol.ac.uk/education.

Burrow, M., Evdorides, H., Hallam, B., & Freer-Hewish, R. (2005) Developing formative assessments for postgraduate students in engineering. *European Journal of Engineering Education*, 30(2), 255-263. doi: 10.1080/03043790500087563.

CCSSO (2008) *Formative Assessment: Examples of Practice*. A work product initiated and led by Caroline Wylie, ETS, for the Formative Assessment for Students and Teachers (FAST) Collaborative. Council of Chief State School Officers: Washington, DC. 2008.

Center for Mathematics, Science and Computer Education: online/ interactive mathematics activities http://www2.eboard.com/eboard/servlet/BoardServlet;jsessionid=6FE830D99D518 952E43C89FBDC402DBC?ACTION=NOTE_SHOW&ACTION_ON=NOTE&OBJECT_ID=2358 08&SITE_NAME=cmsce&BOARD_NAME=mathatthelab&SESSION_ID=em0i3gb2yq19727 9&TAB_ID=81941 [accessed 27-05-13].

Cheung, A. C. K., (2011) The Effectiveness of Educational Technology. Applications for Advancing Mathematics Achievement in K-12 Classrooms: A Mets-Analysis Best Evidence Encyclopedia (BEE) www. Bestevidence.org [accessed 05-06-13].

Carnegie Learning Cognitive Tutor website Accessed at http://www.carnegielearning.com/specs/cognitive-tutor-overview [accessed 03-06-13].

Clesham, R. (2009), Making it real: Interactive assessment of national curriculum science process skills. *Paper presented at BERA (British Educational Research Association)*.

Cook, J., and Crabb, N. (2002) Genuine thinking or random button pressing? Self-assessment design in computer-based learning materials. Winds of change in the sea of learning. *Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE): 8-11 December 2002*, 2002 (ISBN: 04730*91194).

Crisp, G. (2010) Interactive e-assessment – practical approaches to constructing more sophisticated online tasks. *Journal of Learning Design*, 3(3), 1-10.

Crisp, V., & Ward, C. (2008) The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers and Education*, 50(4), 1509-1526. doi: 10.1016/j.compedu.2007.02.004.

CRLI (2009) Center for Research on Learning and Instruction, University of Szeged (2009) (website) accessed at www.edu.u-szeged.hu/~csapo/irodalom/DIA/Diagnostic_Asessment_Project.pdf [accessed 06-06-13].

Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010) Exploring formative e-assessment: using case stories and design patterns. *Assessment and Evaluation in Higher Education*, 35(5), 619-636.

Davis, M., McKimm, J., & Forrest, K. (2013) E-Assessment – The use of Computer–Based Technologies in Assessment: How to Assess Doctors and Health Professionals (pp. 41-52): John Wiley and Sons, Ltd.

Delen, I., Wan-Tzu Lo, Kuhn, A., McGee, S., Duck, J., Quintana, C., Krajcik, J. (2012) Zyde-co: A New Mobile Application Support Scientific Explanations, The University of Michigan, Michigan State University and The Learning Partnership. www.ilearntechnology.files.wordpress.com/2012/12/delen-et-al-2013.pdf [accessed 02-01-14].

Dennick, R. E. G., Wilkinson, S., & Purcell, N. (2009) Online eAssessment: AMEE Guide No. 39. *Medical Teacher*, 31(3), 192-206.

Doorey, N. A. (2012) Coming Soon: A New Generation of Assessments. *Educational Leadership*, 70(4), 28-34.

Feldman, A., & Capobianco, B. M. (2008). Teacher Learning of Technology Enhanced Formative Assessment. *Journal of Science Education & Technology*, 17(1), 82-99.

Furse E., (2009) Computer Based Revision *Proceedings of the BSRLM (British Society for Research in Learning Mathematics)* Joubert, M. (Ed.) Vol.29,no.3: Nov 2009, p25-30.

Gee J.P., Schaffer D.W (2010) Looking Where the Light is Bad: Video Games and the Future of Assessment Edge: the latest information for the education practitioner 6(1), pp3-19 http://edgaps.org/gaps/wp-content/uploads/EDge-Light.pdf [accessed 03-09-13].

Hughes, S. (2006) Assessing mathematics onscreen: What are we assessing? Proceedings of the *British Educational Research Association Conference, Warwick University, September 2006*.

Hughes, S. and Rose, P. (2006) The Computer Mediated Test Project. Report on Phase 1: The development and evaluation of key stage 2 and key stage 3 computer mediated mathematics test items. Research Team, Test Development Team. *Internal report, Edexcel, Pearson.*

Hughes, S., Custodio, I. Sweiry. E. and Clesham, R. (2011) Beyond multiple choice: Do e-assessment and mathematics add up? *Proceedings of the Association for Educational Assessment Conference 2011, Queen's University, Belfast. November 2011.*

JISC (2007) Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education (online) http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf [accessed 06-06-13].

Johnson J. (ed) (2013) Three Things Game Designers Need to Know about Assessment ETS Research Spotlight, Issue 8, April 2013 www.ets.org [accessed 28-05-13].

Kennewell, S (2008) Interactivity in the classroom and its impact on learning. Paper presented at the *British Educational Research Association Conference, Herriot Watt University, September 2008*.

Kennwell, S (2001) Using affordances and constraints to evaluate the use of information and communications technology in teaching and learning. *Journal of Information Technology in Teacher Education*, Vol 10, nos 1 & 2 p 101 – 116.

Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Martin, F., Davies, D., Pollitt, A. and Whitehouse, G.   (2009) E-scape portfolio assessment phase 3 report. http://www.gold.ac.uk/media/e-scape_phase3_report.pdf [accessed 29-05-13] .

Llewellyn, D. (2007) Inquire within: implementing inquiry-based science standards in grades 3-8. Corwin Pr.

McKinsey and Company Transforming Learning through mEducation . Available http://www.mckinsey.com/Search.aspx?q=formative%20assessment [accessed 12-05-13].

Meadows, M. and Billington, L. (2005) A Review of the Literature on Marking Reliability. Manchester: AQA   [online]   available   at: https://orderline.education.gov.uk/gempdf/1849625344/QCDA104983_review_of_t he _literature_on_marking_reliability.pdf [accessed 26-06-13].

Miller, T. (2009) Formative computer-based assessment in higher education: the effective-ness of feedback in supporting student learning. Assessment & *Evaluation in Higher Education*, 34(2), 181-192. doi: 10.1080/02602930801956075.

Neue, F. Enhancements of the Perception™ Assessment Server for Metacognitive Anal-yses, Neue Technologien und Lernen in Europa e.V. (fred.neumann@belab.de) PepCAA www.pepcaa.odl.org [accessed 15-05-13].

Neumann, D. L. (2010) Using Interactive Simulations in Assessment: The Use of Computer-Based Interactive Simulations In The Assessment Of Statistical Concepts. *International Journal for Technology in Mathematics Education*, 17(1), 43-51.

NRich website www.nrich.maths.org [accessed 06-06-13].

Pearson (2013) Abacus Evolve mathematics programme (Pearson) for the primary sector (website)   available   at: http://www.pearsonschoolsandfecolleges.co.uk/Primary/Mathematics/AllMathema ticsresources/AbacusEvolveFrameworkEdition/Structure/Structure.aspx [accessed 24-06-13].

Peat, M., & Franklin, S. (2002) Supporting student learning: the use of computer–based formative assessment modules. *British Journal of Educational Technology*, 33(5), 515-523.

Pellegrino, J. W., & Quellmalz, E. S. (2010) Perspectives on the Integration of Technology and Assessment. *Journal of Research on Technology in Education*, 43(2), 119-134.

Purvis, A. J., Aspden, L. J., Bannister, P. W., & Helm, P. A. (2011) Assessment strategies to support higher level learning in blended delivery. *Innovations in Education & Teach-ing International*, 48(1), 91-100. doi: 10.1080/14703297.2010.543767.

QCDA   (2011).   Key   stage   3   ICT   onscreen   assessment   tasks.   Online   at http://www.qcda.gov.uk/assessment/6555.aspx [accessed May 20, 2011].

Quellmalz E. S., Pellegrino J. W. Technology and Testing (January 2009) www.sciencemag.org Vol. 323 [accessed 02-06-13].

Quellmalz, E. S., and Pellegrino, J. W. (2009) Technology and Testing. *Science*, 323(5910), 75-79.

Ramadoss, R., and Wang, Q. (2012) Evaluation of a web-based assessment tool for learn-ing grammar at the primary school level. *International Journal of Continuing Engi-neering Education & Lifelong Learning*, 22(3/4), 175-184.

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

Richardson, M., Baird, J-A., Ridgway, J, Ripley, M., Shorrocks-Taylor, D. and Swan, M. (2002) Challenging minds? Students' perceptions of computer-based World Class Tests of problem solving. *Computers in Human Behaviour* 18 pp633-649.

Sainsbury, M., & Benton, T. (2011) Designing a formative e-assessment: Latent class analysis of early reading skills. *British Journal of Educational Technology*, 42(3), 500-514. doi: 10.1111/j.1467-8535.2009.01044.x.

SAM Learning www.samlearning.com [accessed 06-06-13].

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., and Irvin, P. S. (2011) Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, 48(9), 1050-1078. doi: 10.1002/tea.20437.

SCHOLAR in Scotland at the school/university interface www.scholar.hw.ac.uk [accessed 07-06-13].

Shermis, M. D., Burstein, J., and Leacock, C. (2006) 'Applications of computers in assessment and analysis of writing' in C. A. McArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 403-416). New York, NY: Guilford Press.

Sluijsmans, D. M., Prins, F. J., and Martens, R. L. (2006) The Design of Competency-Based Performance Assessment in E-Learning. *Learning Environments Research*, 9(1), 45-66. doi: 10.1007/s10984-005-9003-3.

Stödberg, U. (2012) A research review of e-assessment. *Assessment and Evaluation in Higher Education*, 37(5), 591-604. doi: 10.1080/02602938.2011.557496.

Swithenby S. J., (2006) Screen-Based Assessment: New Directions *HEA Journal* Issue 2, December 2006 http://journals.heacademy.ac.uk/doi/abs/10.11120/ndir.2006.00020023 Thousand Oaks, CA: Corwin Press.

Threlfall, J; Pool, PC; Homer, MS; Swinnerton, BJ Implicit aspects of paper and pencil mathematics assessment that come to light through the use of the computer *Educational Studies in Mathematics*, pp335-348, 2007, Kluwer Academic Publishers, 66(3).

Timmers, C. F., Braber-van den Broek, J., and Van den Berg, S. M. (2013) Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers and Education*, 60(1), 25-31. doi: 10.1016/j.compedu.2012.07.007.

Tippins, N. T. (2011). Overview of Technology-Enhanced Assessments. *Technology-Enhanced Assessment of Talent* (pp. 1-18): Jossey-Bass.

University of Akron (2012) http://www.uakron.edu/education/about-the-college/news-details.dot?newsId=40920394-9e62-415d-b038-15fe2e72a677&pageTitle=Recent%20Headlines&crumbTitle=Man%20and%20%20machine:%20Better%20writers,%20better%20grades [accessed June 17 2013].

Van der Kleij, F. M., Eggen, T. J. H. M., Timmers C.F., et al. (2012) Effects of feedback in a computer-based assessment for learning, 263-272. *Computers and Education* 58 (1).

Voelkel, S. (2013). Combining the formative with the summative: the development of a two-stage online test to encourage engagement and provide personal feedback in large classes. *Research in Learning Technology* 21.

Wang, K. H., Wang, T. H., Wang, W. L., and Huang, S. C. (2006) Learning styles and formative assessment strategy: enhancing student achievement in Web-based learning.

Journal of Computer Assisted Learning, 22(3), 207-217. doi: 10.1111/j.1365-2729.2006.00166.x.

Wang, T. H. (2007) What strategies are effective for formative assessment in an e-learning environment? Journal of Computer Assisted Learning, 23(3), 171-186. doi: 10.1111/j.1365-2729.2006.00211.x.

Wesiak, G., Al-Smadi, M., Höfler, M., and Gütl, C. (2013) Assessment for Complex Learning Resources. International Journal of Emerging Technologies in Learning, 52-61. doi: 10.3991/ijet.v8iS1.2354.

Whitelock, D. (2006) Roadmap for e-assessment. Report for JISC, Open University http://www.jisc.ac.uk/search.aspx?keywords=roadmap%20for%20e-assess-ment&filter=s&type=adv&sort=relevance&method=all&collection=default_collection [accessed 06-06-13].

Wylie, E. C. & Ciofalo, J. F. (2008) Supporting teachers' use of individual diagnostic items. Teachers College Record http://www.tcrecord.org/content.asp?contentid=15363 [accessed 13-06-13].

Wylie, E. C., & Dolan, R. P. (2013) The Role of Formalized Tools in Formative Assessment. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Zakrzewski, S., & Steven, C. (2003) Computer-based assessment: quality assurance issues, the hub of the wheel. Assessment and Evaluation in Higher Education, 28(6), 609.

6 Appendices

6.1 Appendix 1 Search terms used in literature review

e-assessment	Computer Based Assessment OR E-assessment OR E-learning OR Integrated Learning Systems OR Technology enhanced assessment
e-learning	Learning analytics OR Intelligent tutoring OR Intelligent measurement OR Digital learning OR Digital objects
Formative assessment	Continuous measurement OR Embedded assessment OR Formative assessment OR Integrated assessment
Summative assessment	Summative Assessment OR Assessment
Inquiry based education	Inquiry based learning OR inquiry OR collaborative learning OR discovery learning OR cooperative learning OR constructivist teaching OR problem based learning OR Inquiry OR didactical engineering OR didactical learning OR didactical situations OR open approach OR problem based learning OR problem centred learning OR realistic mathematics education OR argumentation OR design OR project based learning
Competences	21st century skills OR Competence-based assessment OR Competency based learning OR Key competences
Subjects	STM OR STEM
	Mathematics OR Maths OR Math
	Science OR Physics OR Biology OR Chemistry
	Technology OR information communication technology OR information technology OR Computing

6.2   Appendix 2: Journals searched

Table showing the number of articles found in the targeted journals

Target Journal	Relevant articles found
British Journal of Educational Technology	2
Computer-Based Testing	0
Computers and Education	4
Education and Information Technologies	0
Educational Technology, Research and Assessment	1
European Journal of Education: special issue – ICT and Education	1
Frontiers in Artificial Intelligence and Information and Communication Technologies	0
International Encyclopaedia of Education (Technology and Learning - assessment)	0
International Journal of Computer-Supported Collaborative Learning	0
International Journal of E-assessment (Journal of the E-assessment Association)	3
International Journal of Educational Research	0
Journal of Applied Testing Technology	0
Journal of Computer Assisted Learning	2
Journal of Information Technology in Teacher Education	0
Journal of Research on Computing in Education	1
Journal of Science Education and Technology	3
Journal of Technology, Learning, and Assessment	0
Learning, Media and Technology	0
Research in Learning Technology (Journal of the Association of Learning Technology)	1

6.3 Appendix 3 Sources viewed with no relevant content

Citation/link
Cambridge Assessment website www.cambridgeassessment.org.uk
BECTA (2003) What the research says about barriers to the use of ICT in teaching BECTA ICT Research, Coventry www.becta.org.uk
Inspired by Technology, Driven by Pedagogy. A Systemic Approach to Technology-Based School Innovations. Centre for Educational Research and Innovation, OECD
http://www.mckinsey.com/Search.aspx?q=formative%20assessment Santiago, P., McGregor, I., Nusche, D., Ravela, P. and Toledo D. (2012) OECD Reviews of Evaluation and Assessment in Education: MEXICO
http://isolveit.cast.org/home iSolveIt mathematics puzzles to develop logic and reasoning skills. iPad based apps with very limited focus not related to curriculum, more like puzzles e.g. sudoku.
TAO (Testing Assiste par Ordinateur) http://www.taotesting.com/ Open source e-testing platform from MCQs to simulations. Not relevant for IBE.
Maths.org. No relevant content (except link to NRich which has been explored).
Interactive Teaching and ICT, Swansea Metropolitan University. More concerned with ICT and interactive whiteboards than our remit.
International E-learning Association
Computer Aided Learning Conference
OECD Innovative Learning Environments Project 2010
Quest Atlantis – great resource (educational tasks in gaming environment) but relates to teaching and learning and not to assessment.
MacArthur Foundation http://www.macfound.org/programs/learning/ focus on learning not

ASSIST

assessment
e-asTTle project, New Zealand. Mathematics online assessments: on-screen versions of paper tests. Teacher chooses level, length curriculum strands, 1st stage's level chosen by teacher, 2nd stage adaptive, MCQs and short answer Qs.
Journal of Research in Science Teaching: vol 48, pp 1050-1078 Student Learning in Science Simulations: Design features that promote learning gains. Focus on virtual laboratories and science-simulation software with no reference to assessment or IBE.
Doorey, A. How 2 common core assessment consortia were created and how they compare. (December 2012/January 2013) *Educational Leadership*. Sets out plans for project, no new information for this project.
BECTA: Closed in 2011.
Udacity www.udacity.com: free interactive courses (lots re computer science and mathematics) Comprise short video lectures plus integrated quizzes (non-adaptive). Virtual field-trips, forums with peers. More focused on the learning than the e-assessment model.

The EU project Assess Inquiry in Science, Technology and Mathematics Education (ASSIST-ME) investigates formative and summative assessment methods to support and improve inquiry-based approaches in European science, technology and mathematics (STM) education.

In the first step of the project, a literature review was conducted in order to gather information about the current state of the art in formative and summative assessment in inquiry-based education (IBE) in STM. This report is the output of Pearsons work on this with the purpose of provide a literature review that will inform the development of digital assessments which are relevant to the aims of ASSIST-ME. The objectives of the report is to through the literature identify relevant theories and models, identify strategies used in the evaluation of the models which could inform good practice, identify existing relevant digital assessments, and to identify implications for the development of the digital assessments relevant to the aims of ASSIST-ME.

ASSISTME