1 Developments in On-Screen Assessment Design for Examinations

About the Presenter

Keith Myers is Managing Director of BTL. Keith is a qualified accountant, having trained with Baker Rook (now Baker Tilly) and the spent two years producing accounts and introducing early PC systems in financial services with Provident Financial Group. He followed this with a two year stint working in Bermuda in the off shore insurance industry, and joined BTL having worked as Deputy Finance Director at Harry Ramsdens. He joined BTL in 2000, becoming Managing Director in November 2001, Managing Director. As a qualified PRINCE2 Practitioner, Keith oversees all the day-to-day operations of the business' production unit.

Abstract

This paper draws on examples from e-Assessment development and delivery projects undertaken for a range of UK agencies, including the regulators from each of the 4 nations (QCA, ACCAC, SQA and CCEA), and Awarding Bodies such as Edexcel and the British Computer Society.

The work includes the use of:

- rich media (exploring how video, audio, animation and imaging affect assessment performance, including for candidates with disabilities),

- interactivity and adaptivity (exploring how requiring students to make interactive responses affects achievement and engagement),

- advanced computer-marking techniques (work to mark candidates' prose, mathematical workings, and process as well as output),

- item banking complex items to allow "when ready" assessment, and comparability issues with more traditional assessments.

The paper also discusses how "when-ready" e-assessment is blurring the traditionally clear boundary between summative and formative assessment, and the opportunities open to qualification providers to reshape their assessment offerings to act as learning resources.

Finally we look at how the introduction of ICT to the assessment process is also encouraging Awarding Bodies to reshape their relationships with their customers and investigate new ways of providing products and services.

About BTL Group Ltd (www.btl.com)

BTL (www.btl.com) is a leading UK supplier of technology solutions for e-learning and e-assessment. In our e-learning developments, we provide a turnkey service for the design, scripting and production of learning packages, including components such as needs analysis, assessment, portfolio kits, courseware and accreditation tools. In e-assessment we provide both the on-screen assessment content, and the delivery systems and services to Government Agencies and Awarding Bodies for use in both learning and examination settings This year we are launching our award-winning assessment content development system, CP3, which allows awarding bodies to develop and manage their own on-screen interactive assessment content.

Our UK customers for e-learning and e-assessment include DFES, DWP, QCA, BECTA, BBC, learndirect, RM plc, Edexcel and Pearson, OCR, the British Computer Society, SQA and the Teacher Training Agency.

BTL is independently owned and based in Saltaire (nr Leeds and Manchester), England. We employ approximately 75 staff. Our sister company, Virtual College (www.virtual-college.co.uk/), based in Ilkley, provides e-learning delivery services to industry in vocational and professional areas.

One of BTL's products described in this paper – CP3 recently won 2 awards at the British Computer Society Technology Awards. CP3's lead developer, Andrew McAnulla, won Young IT Practitioner of the Year Award, and the product itself was a medallist in the Best Products of 2005 - Service Products category.

The SQA Solar Project

SQA is an executive non-departmental public body sponsored by the Scottish Executive Education Department. It is the national body in Scotland responsible for the development, accreditation, assessment and certification of qualifications other than degrees. It is primarily funded through qualification entry charges and has an annual turnover of approximately £51m. It employs approximately 650 staff in Glasgow and Dalkeith and there are approximately 1,750 centres approved to offer our range of qualifications, including international centres.

The SOLAR Project (**S**cottish **O**n**L**ine **A**ssessment **R**esources) is funded by the Scottish Further Education Funding Council and is supporting the delivery of HN (Higher National) Qualifications. These qualifications consist of units which are traditionally assessed internally within colleges, followed by an external summative end-of-course assessment.

This is a well-established system and has many advantages, however marking pressures on tutors (who have to mark unit end assessments) coupled with consistency and quality issues with internally set and marked unit assessments discovered during post-hoc verification (which could then lead to unexpected results in the summative tests) meant that SQA considered some possible improvements. We believe these improvements not only offer significant benefits to the community of learners and teachers involved, but they also illustrate the powerful beneficial effect that "next generation e-assessment systems" can have on Awarding Body relations with their customer centres, learners and tutors.

The project set out to provide a community-developed solution to the problem. Tutors in centres were invited to form "subject groups" with the strongest centres in each subject area taking the lead. These groups of tutors were then provided with technology and training which allowed them to develop on-screen objective unit assessments for the HN programme. These assessments are then submitted to SQA for Quality Assurance, before being signed off as live assessments. Centres (including those that authored the tests, and all the other Scottish FE colleges) then can provide these tests online to their candidature. The tests are electronically marked and results are available immediately. In addition, by pre-approving the tests, centres can offer them with confidence from the start of a course, with no risk of problems post-hoc with the validation.
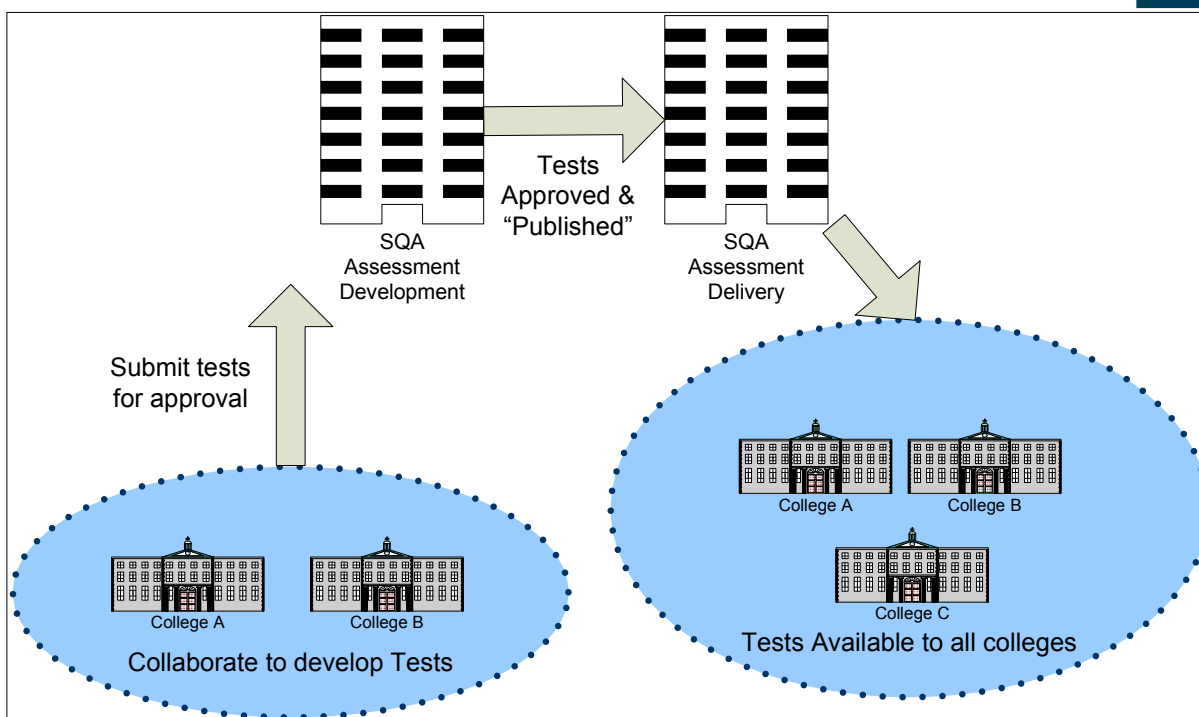
Figure 1 – Outline Process

Experiences in the Project

Broadly the project has been a success - it is now entering its 3rd phase, with approximately 50 colleges using 320 tests supplied by a community of 40 authors. By the end of the project we expect to have nearly 700 live tests on the system.

Throughout the programme, the implementation of the technology has caused considerably less problems than human factors – mainly communication and training. This is counter to what many expect to find – i.e. that the technology is now stable, but requires considerable skill in both using it, and applying it within the organisations. This has been particularly the case for the assessment development, where considerable training on both technical and educational (assessment design) aspects was required.

SQA and BTL's findings in the project are as follows:

- Training session on using CP3 authoring system and in assessment design is a constant and ongoing requirement – training at the outset is unlikely to be sufficient. The additional factor of multiple author communities in multiple locations, with multiple abilities brings multiple challenges.

- Customers and suppliers need a common understanding of project expectations and priorities.

- Success within the project was more about the suitability of the curriculum than technology (which broadly delivers as promised)

- There is no single eAssessment system that can provide all that a Qualification Authority requires

- Essential to adapt requirements based on user experience, and to work particularly hard on communication between all parties at all times. This has implications in terms of support and project management.

- The Invitation-To-Tender procurement process is problematic where the project concerned has evolving requirements (due to both lack of certainty at the outset and the inevitable experiences gained from running a highly innovative project).

- Having made these points, the experience of the authors concerned has been ultimately positive in that they believe they have learnt about e-Assessment, assessment design, and about their own subject area - an unexpected benefit of the project.

Supporting Innovative Assessment Delivery

As a supplier of exam systems recognised for their ability to support innovative assessments (both in terms of the content, and the delivery modes) BTL was interested in the SQA project because it offered the possibility of connecting development and deployment systems in a web-enabled setting.

In the UK, our experience is that first generation e-assessment projects generally start with replication of existing paper processes (this applies to both the test development and test delivery phases). In addition to the obvious familiarity benefits of this (and therefore reduced risk in the technology requirements specification process) there are also advantages in terms of proving the comparability with paper tests, which often continue to run in parallel.

In subsequent phases, organisations begin to explore the specific benefits of on-screen assessments (in terms of efficiency and effectiveness gains). These are well documented in other projects but can include:

- Flexibility of delivery in terms of time, pace, place

- Immediate results: in addition to allowing rapid progression this can help bridge the traditional gap between formative and summative assessment: By providing tutors with immediate (and therefore useful) feedback about the detail of learners' performance in specific areas.

- Operational cost savings in centre.

- Supporting institutional objectives of leveraging use of ICT.

- Providing more valid assessments by assessing a broader range of skills/knowledge in more realistic settings.

In projects such as the SOLAR programme, although there are significant benefits from moving to on-screen development and delivery, the UK experience is that there is no desire to compromise on areas of assessment that have been seen as traditionally important. For example, the move from human-marked to objective computer-marked assessments is treated with careful scrutiny, and the introduction of computers brings an expectation among teachers and learners alike that the on-screen assessment will make good use of the interactive and rich media capabilities of modern computers.

BTL saw the critical technology requirements of 2nd generation assessment systems are as follows:

- Providing a **distributed test development process** that supports workflow among a community of people with different roles and skills.

- **Need to deploy development and delivery tools across an entire assessment enterprise** – becoming less of a project and more of a mainstream activity (although paper systems often continue in parallel, of course)

- Support for the key benefits of ICT in assessment:

- o **Rich media**.   Self-evidently, computers can deliver a wider range of media types than paper.   Most notable are the following: animation and video (with play, pause, slow motion and replay), audio, and use of colour. Simple use can lead to significant validity improvement: e.g. much of the UK literacy curriculum is about observing and participating in face-to-face and telephone interaction with others. Paper is weak at conveying such scenarios with good face validity: the simple use of video and audio adds greatly to the validity.

- o **Interactivity**. Interactivity is useful primarily in two ways. Firstly it allows candidates to give answers to more complex questions without necessarily having to write their responses down in text. Secondly it offers the opportunity for simulation systems. The ability of a learner to observe a system, manipulate some of its parameters, take further observations, draw hypotheses and test them out, etc. is a crucial feature of many curricula and is well-supported by on-screen interactive content.

- o **Adaptivity**. As a subset of interactivity, the ability of a system to adapt to its users activities is of great interest in assessment. This can speed up assessment and also provide increased motivation for learners in formative settings.

- o Advanced computer marking. Using advanced computer techniques to improve the range of assessments that can be marked electronically (for example, marking diagrams, free text, mathematical formulae and processes).

- • **Powerful Item Banking** – to support the ongoing development of new items, and modification of existing items in a bank while the bank is also being used to generate live assessment content.

- • **Supporting Formative Assessment alongside Summative Assessment**.

Alongside these benefits of on-screen assessment are significant, they bring potential problems which development and delivery systems must seek to deal with:

- • Complexity, cost of development & trialling can increase

- • Issues with accessibility for learners with disabilities may increase

- • Technical deployment may become more challenging, for example raising the minimum system specifications for PCs, servers or network bandwidth.

- • More learner and teacher preparation may be necessary to ensure students are aware of what they are expected to do, and how to operate the ICT properly in order to do it.

The outline structure of BTL's system is shown in the diagram below:

The presentation which accompanies this paper will elaborate on some of the system's features and how they benefited SQA. The following features are particularly worthy of note:

- • Item development ranges from the very simple to the very powerful. The development platform uses templates to allow rapid and simple creation of basic items, but leverages the full powers of Flash and XML to support more complex items, tests and curriculum taxonomies.

- • The development phase is abstracted from the final delivery platform, allowing content to be produced and then published to a variety of output forms at a later date. This allows (for example) a bank of traditional items to be held in XML form

and output to either on-screen or on-paper at the time of test assembly. It also allows practice tests to be published for delivery in other systems (e.g. within a VLE).
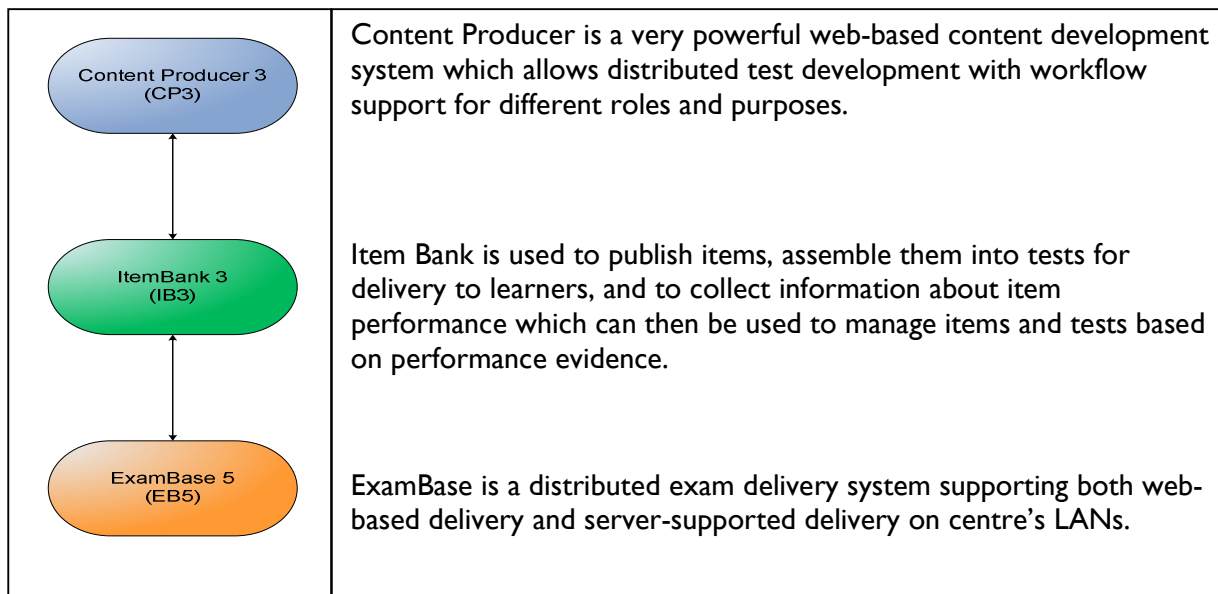
Content Producer 3 (CP3)	Content Producer is a very powerful web-based content development system which allows distributed test development with workflow support for different roles and purposes.
ItemBank 3 (IB3)	Item Bank is used to publish items, assemble them into tests for delivery to learners, and to collect information about item performance which can then be used to manage items and tests based on performance evidence.
ExamBase 5 (EB5)	ExamBase is a distributed exam delivery system supporting both web-based delivery and server-supported delivery on centre's LANs.

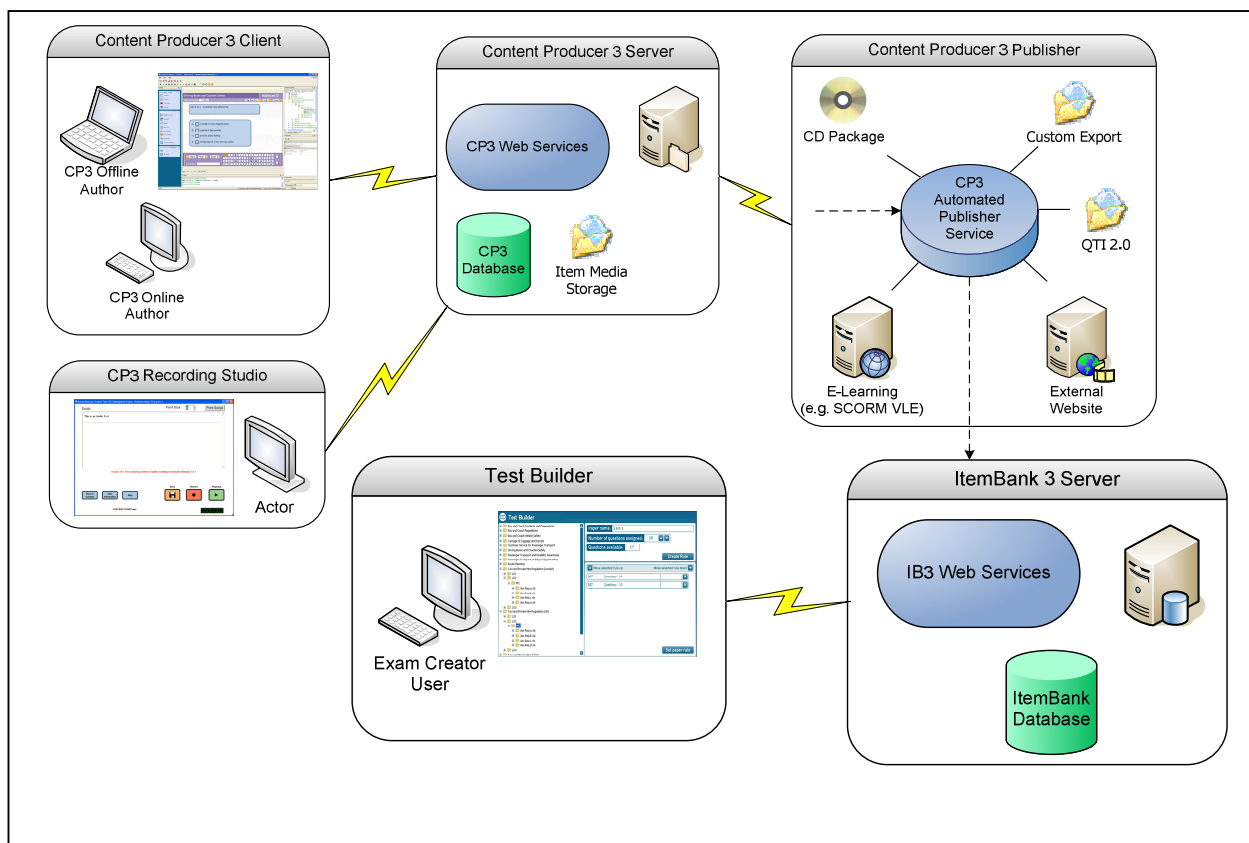*Figure 2 – Core Assessment System Components*

Figure 3 – System Functionality

- There are effectively two item banks. The first, part of the CP3 content production system is for items in development, at various stages in their workflow. These items are free to be edited according to the rules of the workflow and the user's role. Once published to the ItemBank IB3 Database, the item is fixed – potentially being

used in live examinations and having candidate data stored about its performance. Modifications to the item must be made in the content development system and the 'new' item must then be republished.

- The rules for assembling tests (both static and dynamic, i.e. fixed form and containing randomised elements) are highly complex, and subject to user control. Considerable effort has been devoted to producing a user interface for this test construction process which is sufficiently powerful but simple enough to be used by a Subject Officer to manage an examination.
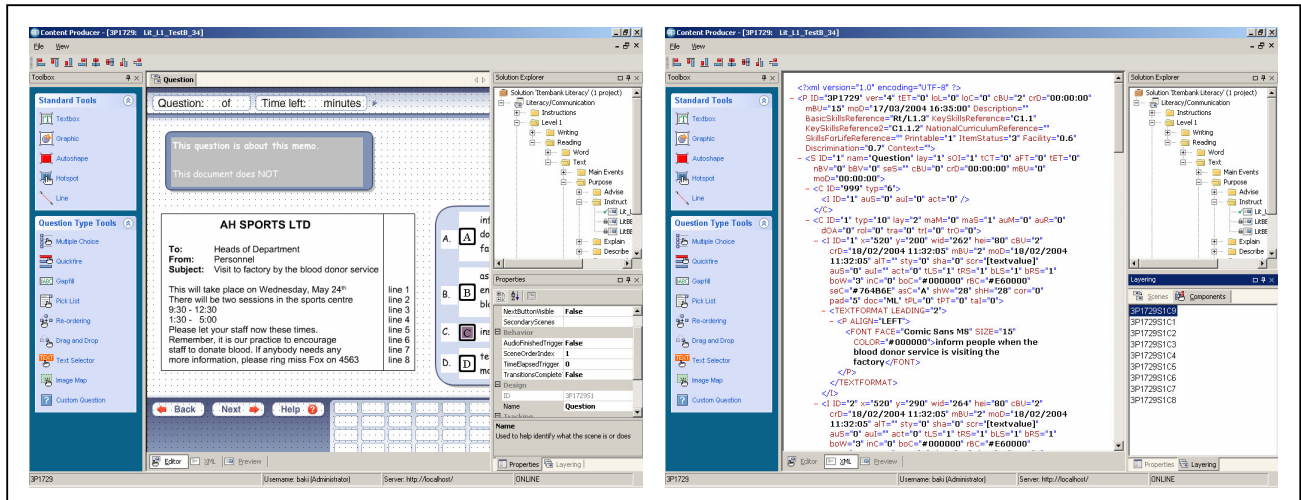

Figure 4 – Examples of CP3 Development screens showing XML and WYSIWIG Views

The CP3 development system is supported by a substantial team of developers and used by BTL's in-house production team for client content development (in fact the same system is used for e-learning and e-assessment content). However in deploying the system in customer centres (e.g. Awarding Bodies) to allow in-house content development, the additional supporting features have been required:

- A telephone and email helpdesk offering technical and assessment design support and advice.

- A maintained and supported FAQ and User Guides, including simple "How To" Tutorials for occasional users

- Template playbooks detailing all the (~150) item types that CP3 can support as standard.

- Systematic processes of qualifying trainees as capable to use the system. Currently we operate a 3 tier structure for CP3 producers with access to different features at each level, to ensure that users who are still learning do not stray into areas of "dangerous" functionality. This programme is supported by a series of tests and examinations (and these are used as part of the HR/personnel performance review programme within BTL).

- A carefully managed programme of upgrades. The CP3 system is under continuous development both to meet specific customer requirements (for example recent work includes improved support for accessible content and the ability to import and output QTI IMS v2.0 content). While upgrades for internal staff can be rolled out with informal communication, it is important that upgrades are both planned and notified in advance to avoid external users simply seeing additional or different features on the desktop.

Within the examination delivery system which accompanies CP3 (called ExamBase) we have seen rapid increase in both the volume of centres and the number of tests (the graph below demonstrates take-up on one of our customer's assessment programmes). Alongside this growth, we have seen a corresponding decrease in technical problems with installing new centres which we attribute to a combination of improved process and increasing user readiness for e-assessment.
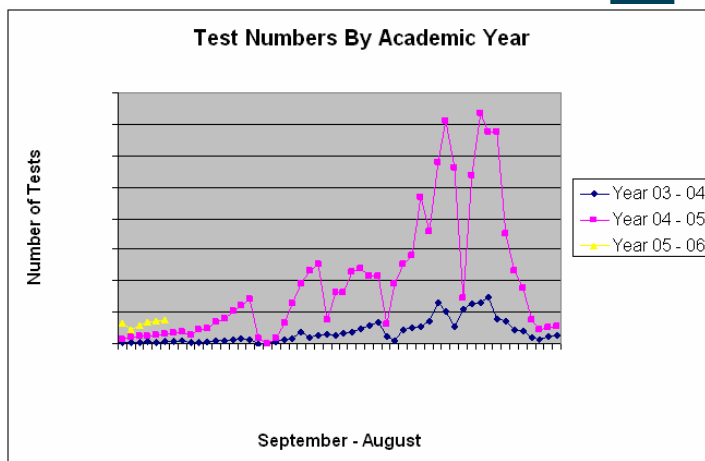

Figure 5 – E-Assessment Take-Up

Developments in Formative Assessment

Considerable work is underway (in parallel with e-examination development and deployment) to use the power of ICT to provide immediate powerful and detailed feedback from formative assessments which can be used as part of the learning process. One example of this is the suite of tools developed for the English Department for Education & Skills (Ministry of Education) for the Skills for Life Qualifications. Formative Assessments exist at each of the interventions in the diagram across.

Supporting the production of on-screen assessments by external authors where feedback frames are included is complex, as the feedback itself is effectively an additional set of conditional screens based on the marked outcomes of the questions. Our presentation will demonstrate recent examples of innovative work in this area.
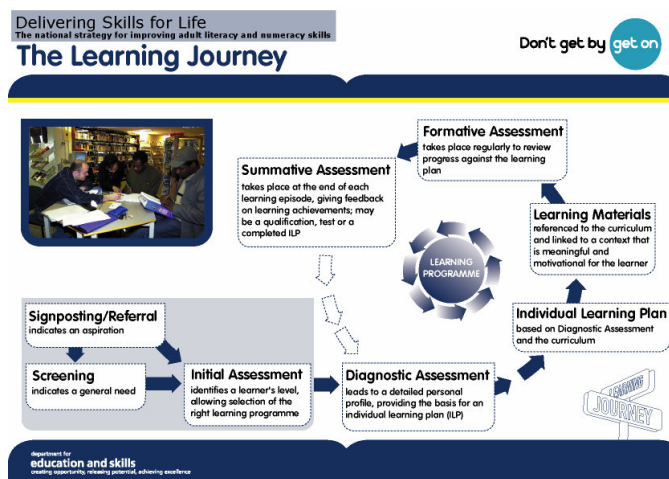

Figure 6 – The Learning Journey

One current view of how best to tackle formative assessment is set out by Black and Wiliam's "Working inside the black box" (Kings College, London), which holds out the promise of very significant achievement gains if the formative assessment techniques are used. However, the impact of this work on mainstream teaching has been limited to date, possibly because the demands it places on individual teachers with large groups of learners are simply too great.

There are a number of levels at which the feedback can take place:

1. It may refer to a group of questions, usually through a mark or a simple qualitative comment following some written responses.
2. It may refer to an individual question, following verbal questioning in a group or on an individual basis, either verbally or on paper.
3. It may refer to one step in a question, with the teacher looking over the shoulder of the learner and pointing out a mistake as it occurs, or marking a question with meticulous care.

All of the above take place in a traditional teaching and learning context, but limitations on teacher time mean that learners get more feedback at level 1 than at level 2, and in turn more at level 2 than at level 3. The opportunity presented by e-learning is to provide much more feedback at level 3, because the computer does not have the limitations on time faced by the teacher.

In our view it is not realistic for the computer to provide feedback at level 3 of the traditional type ("explanation") except in very rudimentary form. This is because the number of possible responses required is vast (it is known as a combinatorial explosion), and cannot be programmed in. "Online Help" systems seem so wooden and stupid because of this problem.

On the other hand, it is much easier to track the learner's work electronically and highlight an error as soon as it occurs. This has the advantage of leaving the learner with the cognitive conflict, an important part of the learning process, and also a clear view of the precise location and nature of the problem. All this adds up to the ideal conditions for learning. Its nearest equivalent is a teacher looking over a learner's shoulder and pointing out a mistake as it occurs – but answering further questions with questions rather than explanations. The computer is ideally situated to deliver at least parts of this kind of Socratic Dialogue.

Our recent work in ICT-supported formative assessment seeks to provide the learner with immediate and relevant feedback at the point of error in order take advantage of both the elements of Wiliam and Black's recommendations regarding Assessment for Learning, and the lessons learned regarding the benefits of immediate results/feedback to learners in terms of achievement and motivation. In addition to helping the learner progress with a problem, advances in ICT-mediated Formative Assessment also hold out promise for classroom teaching - helping teachers to manage the large amount of performance information that the assessment is providing, thereby providing timely information to focus teaching effort.

We hope to present out initial findings from trials of these new assessments at the conference.