ITEM BANKING FOR INNOVATIVE ITEMS AND EXAMINATIONS IN THE UK

Martyn Roads and John Winkley

Item Banking for Innovative Items and Examinations in the UK

Martyn Roads
M R Educational Consultants
Sevington Park
Maidstone
Kent ME15 9SB
miroads@dial.pipex.com

John Winkley
BTL Group Ltd
Salts Wharf
Ashley Lane
Shipley
West Yorkshire BD17 7DB
johnw@btl.com

1 Abstract

BTL, together with a number of partners, have been working with the UK Department for Education and Skills (England) to develop an item banking system to deliver formative assessment material for Key and Basic Skills teachers and learners to use as a learning resource.

The Key and Basic Skills developments are part of the UK Government's *Skills for Life* strategy which is designed to encourage learners of all ages to address their Maths and English skills and to ensure that they try and achieve a Maths or English qualification at at least Level 2. To this end, the Government has set a target of 750,000 learner achievements by 2007 and is supporting the *Skills for Life* programme with a range of both teacher and learner support mechanisms. Until recently, most learners completing a *Skills for Life* programme of learning would have had the opportunity to undertake a Literacy or Numeracy National Test which, although multiple choice, would have been delivered in a traditional paper based format. Over the last 18 months, however, an increasing number of learners have had the opportunity to take the tests using Computer Aided Assessment (CAA). One of these delivery systems was described by Chris Sealey and Paul Humphries in their paper to the 2003 IAEA conference in Manchester. Learners taking the tests using CAA have been able to book the tests on demand, receive the results immediately after the tests and to benefit from a process that has taken advantage of the technology available. The development of computer

delivered items has provided the incentive and the opportunity to develop the item banking project described in this paper.

Phase 1 of the project was completed on the 31st March 2004 and the system made available to teachers and learners. Proposals are now being finalised (as at end April 2004) for Phase 2 of the project.

The first part of this paper provided an overview of the item bank system, the way it interfaces with item authors and the user features.

However, the possible introduction of item banking into the UK examinations system and the potential for its wider use in accredited examinations raises a number of other issues in relation to the changes that would be required in the traditional examination process to take full advantage of item banking and CAA generally. The second part of the paper identifies a number of these issues and the steps that might need to be taken to address them.

The paper concludes that there are many advantages to be gained from the use of item banks as part of the general development of CAA systems. Not only do they provide teachers and learners with a powerful formative assessment tool but they open up significant opportunities to revise the examination processing system. However, if they are to play an appropriate part in the development of high stakes summative CAA assessment systems, there are a number of process and quality assurance issues that will need to be addressed through appropriate research, pilots and trials before they can be adopted on a wide scale. In some cases, these issues are different for an item bank producing items for on-screen use as well as the traditional paper delivery formats.

2 Introduction

The purpose of this paper is to describe the first phase of the creation of a *Skills for Life* formative assessment item bank, the proposals for the second phase and the impact its use will have of teachers and learners. The paper then continues by examining the issues raised by the potential use of an item bank to deliver high stakes, summative assessment.

Key success factors for the first phases are: the ability of the system to deliver positive value for the *Skills for Life* teacher and learner, ease of use by the teacher and learner, ease of use by content creators, i.e. item and test authors and the ability of the system to handle a range of item types including innovative question types. Whilst these success factors are the prime criteria against which success will be judged, the project team are also taking cognisance of the potential use by exam boards to create live tests and of the system to handle item and examinations other than the *Skills for Life* National Tests.

3 Project Background

The “Item Bank” project is an initiative from the Adult Basic Skills Strategy Unit with the Department for Education and Skills. The project was initiated in August 2003, with the aim of completing the first phase by the end of March 2004. BTL were commissioned to undertake it with a number of partners[1]. The main objectives of the Item Bank Project were to:

- Develop an item banking system which could manage the process of authoring, assembling, delivering and reviewing Skills for Life practice tests and formative assessments to support learning
- Enable remote, on line management of the authoring of items and tests
- Enable teacher and, in due course, learners to assemble complete practice tests or smaller topic based tests to support learning
- Collect item data for review and to enable test developers to modify or delete items from the bank, based on their performance
- Demonstrate that, through the use of expert tagging, the system could produce balanced tests
- Demonstrate that the system could manage a variety of item types including those that could:
    - contain various forms of media such as sound and video
    - items that include high levels of interactivity
    - items of varying length/marks
- Produce the tests in a variety of formats (inc. on screen and on paper)

Further phases of the project are likely to address:

- The validity of the test construction process and whether this could lead to the use of the bank for constructing high stakes test for accredited qualifications
- The provision of additional learner support
- The use of the system to construct other types of tests, including:
    - Screening and initial assessment tests
    - Fully formative tests (moving away from solely minimum competence items)
    - Adaptive test components (in the above)
    - Tests for particular settings (e.g. to support particular pieces of learning materials, learner workplace or lifestyle contexts).
    - Closer links to dynamic learning plans. (UK adult education uses learning plans as both the primary achievement and progression log and as the currency of funding arrangements).

[1] The project partners were BTL, Alpha*Plus, The* Hamlet Computer Group and Greystones Publishing

- The ability of the system to deliver items/tests into a variety of CAA delivery systems using appropriate standards and interoperability definitions

This paper is concerned largely with the first phase of the project.

4 The Item Bank System – an Overview

The item bank system is comprised of three main elements shown in Fig. 1 below.

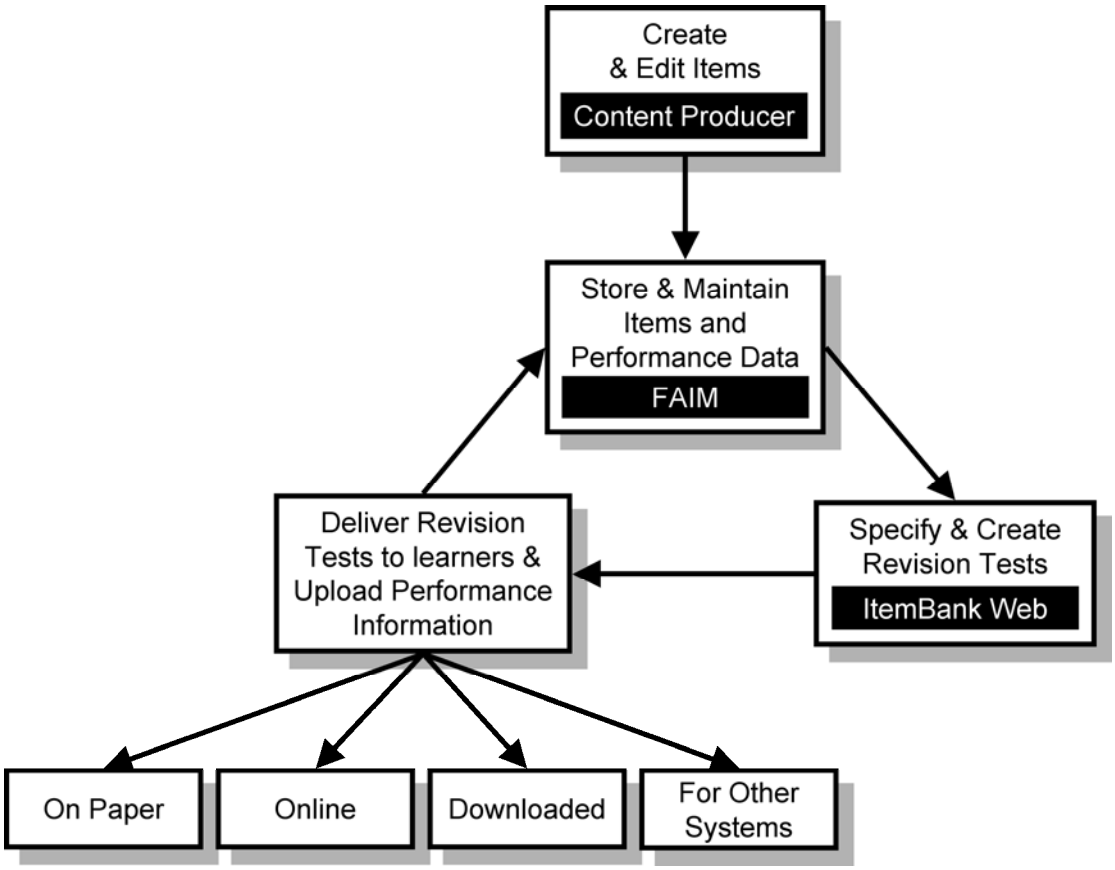


Fig 1 The Item Bank System

1 Content Producer (CP3)

This is a proprietary tool developed by BTL to facilitate the design and production of the items. It can deal with a range of item types and enable the user to see exactly how the item will be rendered on screen or on paper to the learner. At the same time as creating the item, the author provides the tagging detail that allows the item to be stored and then selected against the appropriate criteria. Each item is tagged with a range of information relating to which part of the both the Skills for Life and Key Skills curriculum it matches, which part of the UK National Curriculum is matched, the context of the item scenario, whether it can be printed (or only shown on screen) and its current performance values.

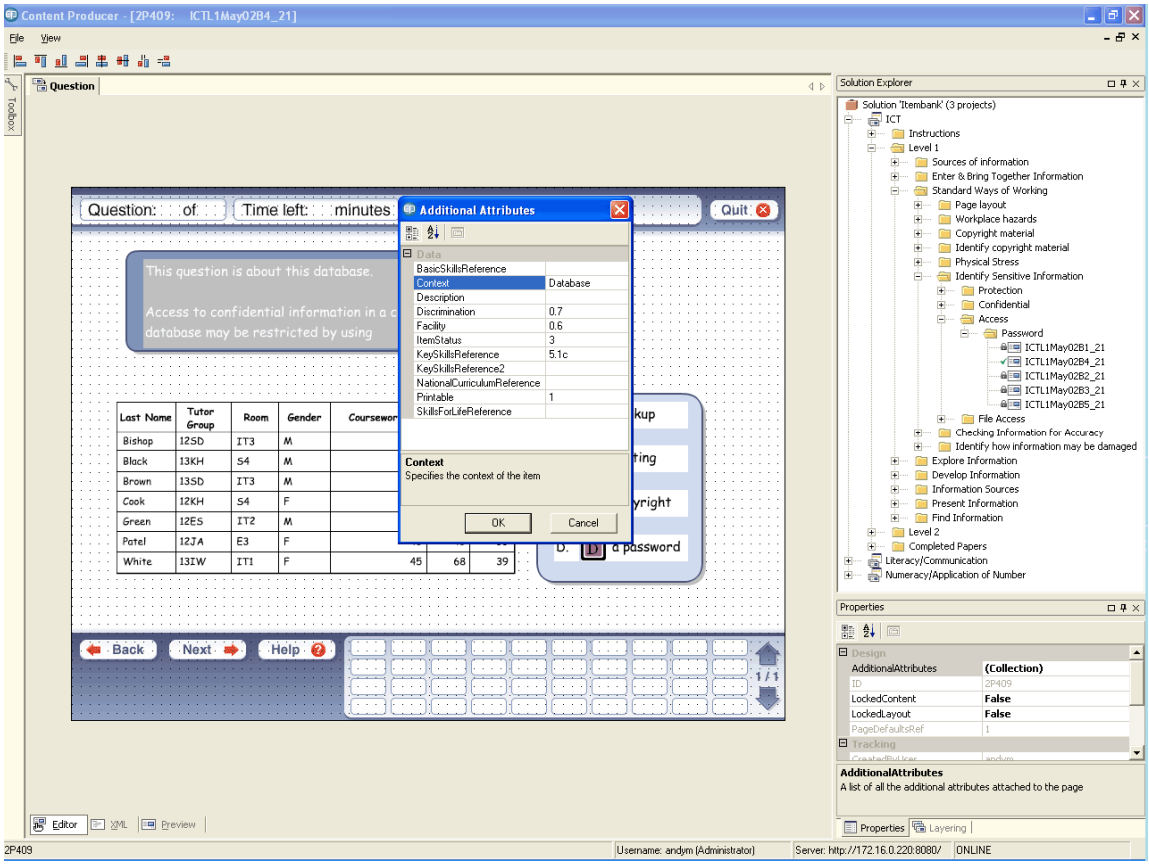


Fig 2 Content Producer showing the attributes Window and the hierarchy of Items with the ICT Subject Area

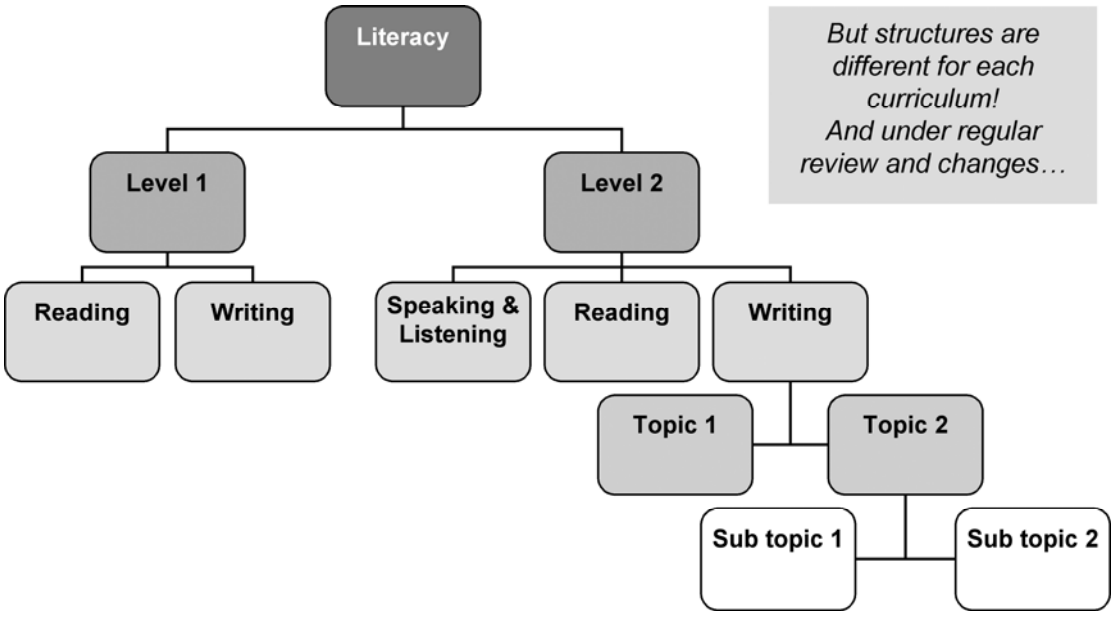
Content Producer is used by BTL for all its e-learning and e-assessment activities – its functionality here is connected directly to the FAIM item bank. The project has proved that connecting a powerful content production system (based largely on production of linear and style-sheet based content) to a bank has considerable benefits for both. The following features of Content Producer add to FAIM's item storage functionality considerably:

- De-skilled content creation. CP3 enables the production of highly engaging learning and assessment content without the need for programming skills (by using templates), thereby reducing production costs.
- *Internet accessibility.* CP3 development tools can be accessed across the internet making it possible for external authors to use the system to create, review and edit content.
- *Greater management control.* CP3 functionality enables workflow management and version control, minimising the risk of production costs spiralling (this is especially important for large projects).
- *Template-based specification of look-and-feel*. Content creators can save time by using templates for sections that require the same look-and-feel.

- *Many activity templates built-in*. A wide range of 'standard' activities can be incorporated using templates, saving much time for content creators.
- *Linking Learning and Assessment*. CP3 is BTL's tool for both e-learning and e-assessment development. This means that combined developments such as formative tools, and assessment for learning projects are easy to develop in a single environment.
- *Mechanisms for building complex learning content*. CP3 enables complex learning screens to be constructed in a standardised way.
- *Ready incorporation of visual assets*. CP3 makes it straightforward for content creators to 'pull-in' graphics, animations and video. This allows the clear distribution of different production activities between different teams with the appropriate skills.
- *Simplified integration of audio content*. CP3 provides advanced facilities for efficiently integrating audio content – a task that has traditionally been complex to manage.
- *Support and training*. BTL supports the CP3 content production process by providing in-house teams of experienced CP3 users and multimedia editors, and through the provision of training for external authors. This gives clients a range of content production options.

2 FAIM

Each item is then stored in a hierarchical structure within the bank using a proprietary system called FAIM. The FAIM software allows the end user to access the item at any level of the hierarchy. Whilst the structures may change from subject to subject and between qualifications the principles remain the same. A typical structure for *Skills for Life* Literacy would look like this.



FAIM also collects and stores the performance data for the item each time the item is used within an online test.

Items are tagged using meta tagging in XML in QTI format where appropriate (although in practice most of the more innovative question types are not supported by IMS QTI in its current form).

Each item in this setting consists of a zip file of content (including media items such as pictures, audio, video, interactive modules, the question itself and mark scheme) with an XML manifest file attached, as shown in the diagram below. A complete examination simply consists of a number of these items zipped together into a single file specified with its own manifest.

3 Delivering Tests

The key to the successful delivery of tests or revision items to the end user is the ability of the end user interface to specify their requirements accurately and for the bank to identify items that both meet those requirements and deliver a balanced test.

The current "live" *Skills for Life* and Key Skills Tests are constructed according to a set of rules that ensures learners receive a balanced test with means each test:

- Contains the same number of items related to each of the principal topic areas
- Contains the same balance of "easier" and "harder" items

- Contains a balance of contexts – scenarios that are work, leisure or home based
- Has a performance which results in pass mark that is within a narrow, set band.

These criteria are achieved in the "live" tests because the tests are constructed by expert test authoring teams who quality assure the tests and the individual items before they are used. The performance of the "live" tests is also monitored, and can be adjusted, through post test moderation and standardisation (see Test Process – Quality Assurance issues in Section 6).

Furthermore the algorithm needed to ensure that in selecting, for example, "spellings" within the Literacy curriculum, the user did not receive 6 items which all reflected the same type of "spelling" issue. Therefore some items are mutually exclusive and the algorithm is constructed to ensure that tests are built appropriately.

Constructing an algorithm that can achieve the same result as expert test developers is one of the most challenging aspects of this project. The use of the tests created for practice use by learners and tutors will provide valuable feedback on the effectiveness of the algorithm and enable it to be fine tuned.

4 Test Delivery

A key objective of the first phase project was to enable tutors and, in due course, learners to create tests for practice use.

The process is managed through a website[2] which can be accessed from any PC with internet access. Whilst a broadband connection speeds the process, it works equally well with a 56 kbps modem.

The whole process needed to be automated and to provide an intuitive user interface that would allow the user to:

- Create practice tests that mirrored, as closely as possible, a live test, complete with a pass mark
- Create shorter tests based round a subject topic
- Have the tests delivered in variety of formats, including live on line, on screen, on paper as a .pdf file and on paper and via a file delivered as an e mail attachment

The user interface consists of a series of screens, each one of which takes the user further down the hierarchy of test and item attributes.

An outline of the screens and the hierarchy is given in Figure 3.

[2] The interface can be accessed at www.itembank.org.uk

Figure 3.

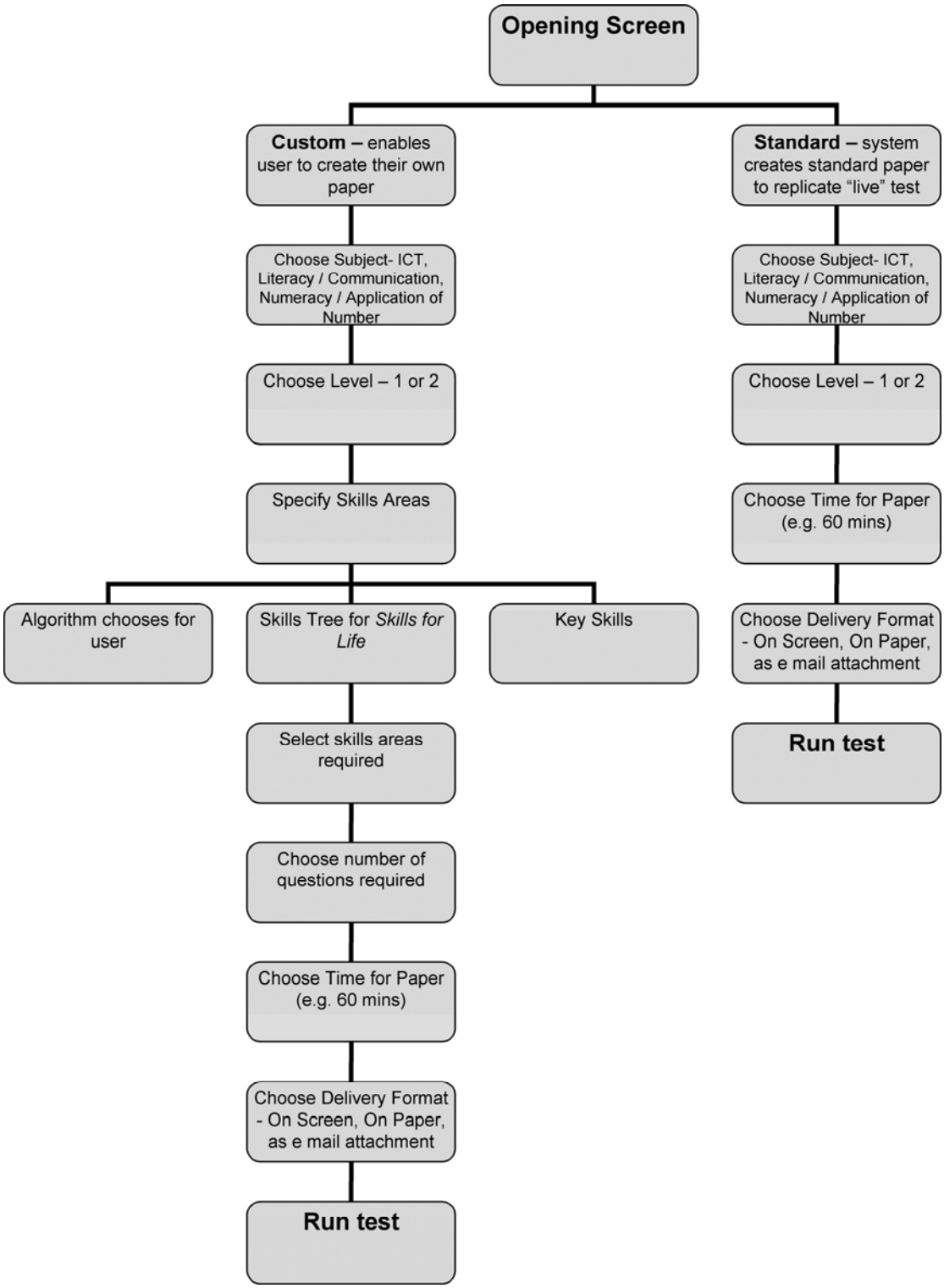
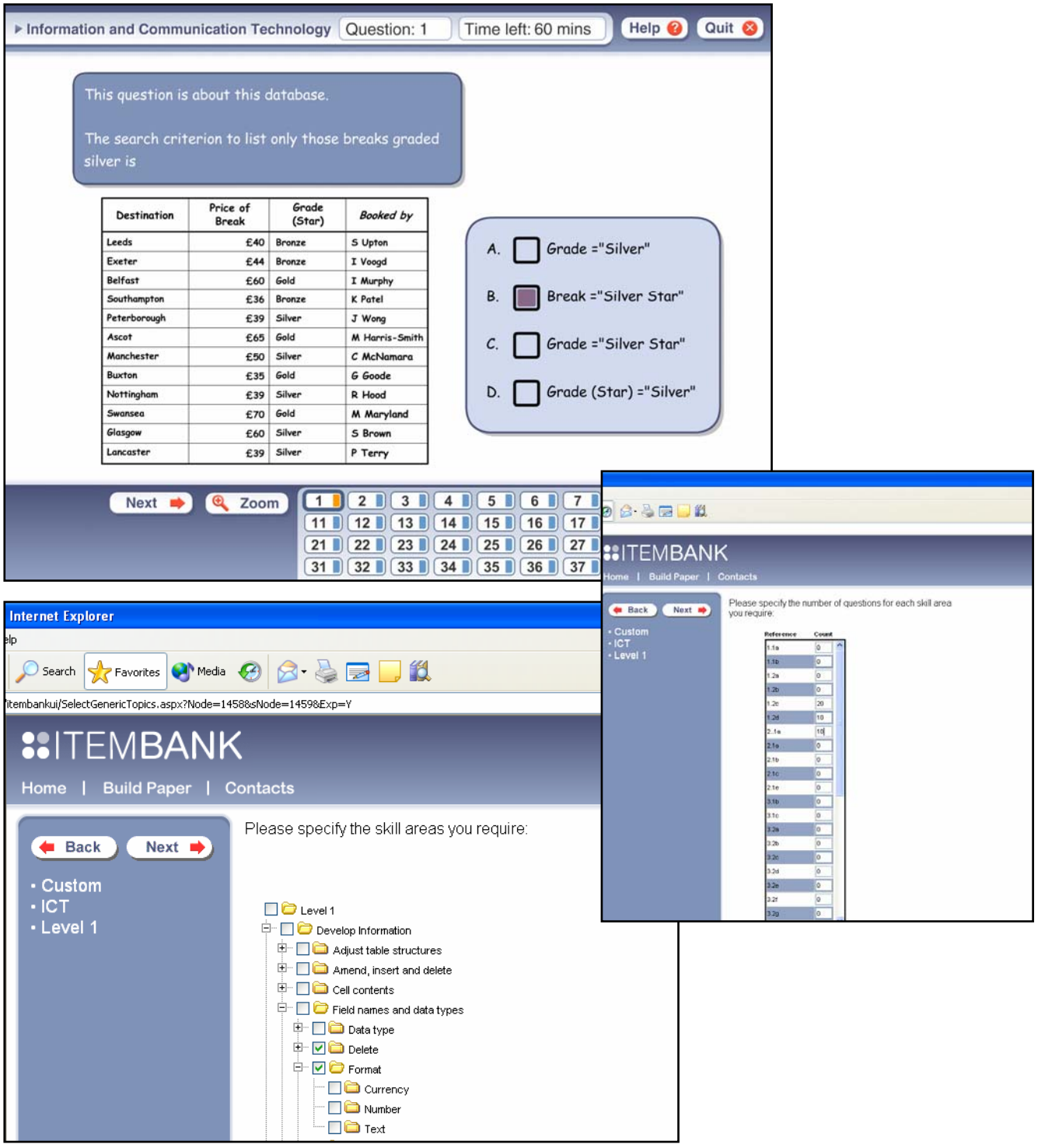

Opening Screen
Custom – enables user to create their own paper
Standard – system creates standard paper to replicate "live" test
Choose Subject- ICT, Literacy / Communication, Numeracy / Application of Number
Choose Subject- ICT, Literacy / Communication, Numeracy / Application of Number
Choose Level – 1 or 2
Choose Level – 1 or 2
Specify Skills Areas
Choose Time for Paper (e.g. 60 mins)
Algorithm chooses for user
Skills Tree for *Skills for Life*
Key Skills
Choose Delivery Format - On Screen, On Paper, as e mail attachment
Select skills areas required
Run test
Choose number of questions required
Choose Time for Paper (e.g. 60 mins)
Choose Delivery Format - On Screen, On Paper, as e mail attachment
Run test

Fig 4. Outline of Screen Hierarchy. An example of the test produced (with some screens from the selection process) is included below:



5 Feedback

In the Phase 1 design, the learner receives feedback showing which items were answered correctly and how items were correct out of the total available. However, the system is also capable of giving both the learner and their teacher much more detailed feedback. Initially this will enable the learner to

identify why a particular item was answered incorrectly, which answer was correct and why. The learner and teacher will also be able to see trends in the types of items learners are finding difficult or the skill areas that are weak. These will be available both at a learner level and at a group level.

6 Items

There are currently 4500 items in the bank, drawn from existing papers and from material authored for "live" papers but not used. All the items are multiple choice and scenario based. As new items become available, they will be added to the bank and older, poor performing items will be removed.

Our trials in these and associated areas have highlighted the following reasons why on-screen questions are preferable to on-paper questions in the basic skills domain (from the point-of-view of both the assessor and improving the assessment):

Richness of 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, use of colour. Of the three areas described in this section, media richness is the easiest to implement and is often regarded as a cosmetic factor. However our work has shown that this underestimates the impact on validity, particularly in areas that draw heavily on "real life".

For example, much of the UK literacy curriculum is about observing and participating in face-to-face and telephone interaction with others. Paper is extremely weak at conveying such scenarios with good face validity, and this would be particularly difficult where cultural variations render the scenarios less familiar to learners. The simple use of video and audio adds greatly to the validity.

Therefore, before even adding any interactivity, the simple use of video and animation to show a phenomenon adds greatly to both the range of questions that can be asked (increasing the assessable area of the curriculum) and the depth of understanding that can be assessed (more challenging questions can be designed around richer media content).

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. We believe that this will help us achieve cultural acceptance, and bridge some of the challenges in designing mark schemes. For example, there are several areas where a candidate may be able to give their answer (or part of it) by manipulating a ***system*** into a particular ***state***, (or sequence of states), or assemble a series of ***components*** into a ***system***.

Secondly interactivity is useful because it offers the opportunity for simulation systems (for example observing the growth of plants with

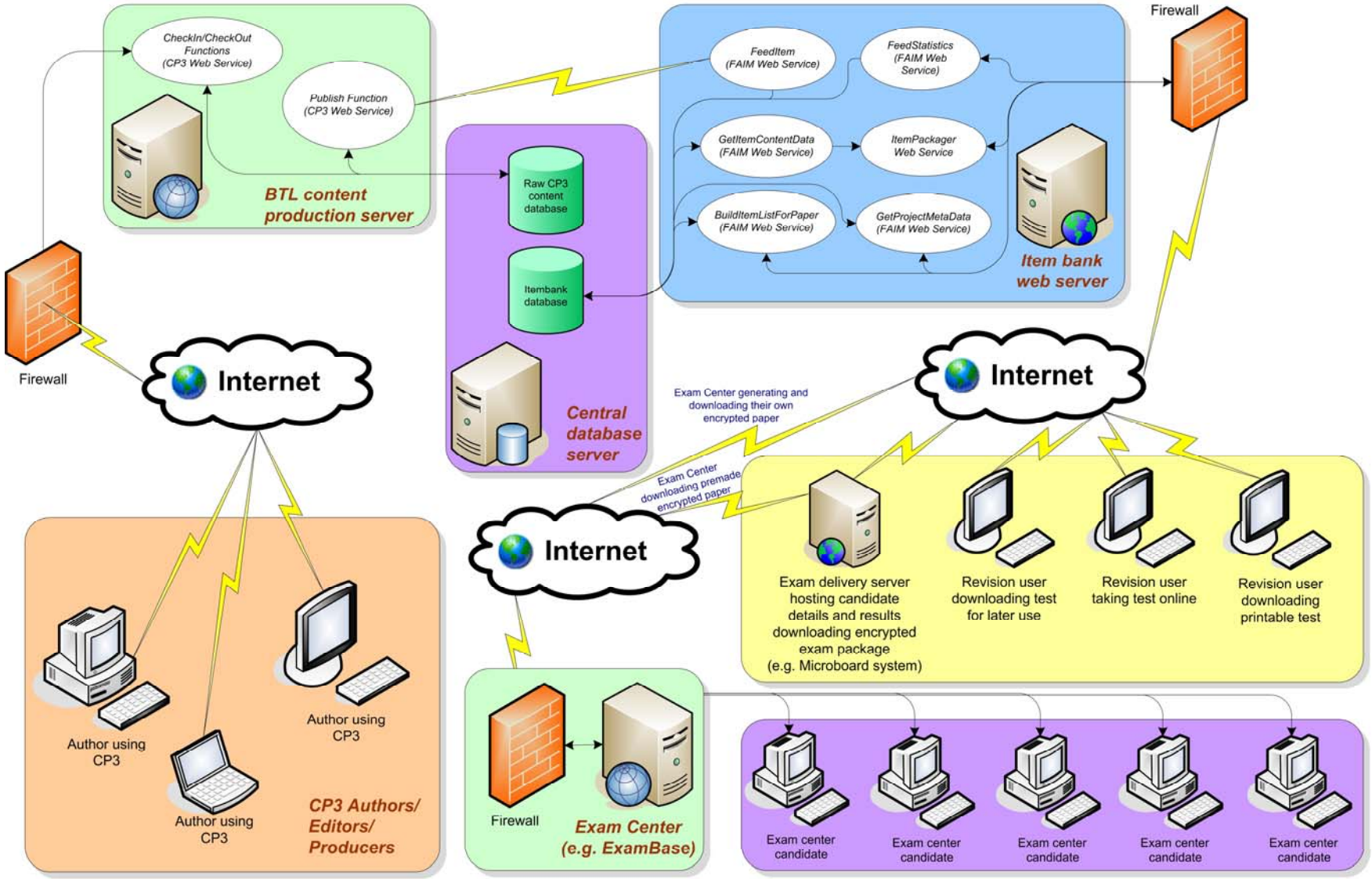
different soil, light and water conditions). The ability of a learner to observe a system, manipulate some of its parameters, take further observations, draw hypotheses and test them out, etc. is a crucial feature of many curricula and is well-supported by on-screen interactive content. The system's performance is repeatable, and the experimental activities are not corrupted by external factors (the computer system can be as pure as required). Essentially, the tests capture a range of response information from test takers over and above dichotomous indications of correctness without as much cultural/language bias as text can produce. This is an area in which BTL has worked for many years and we have learned that the possibilities that ICT-based interactivity offer for assessment designers are so broad as to be overwhelming in some instances. Elsewhere in this proposal we describe our ways of training writers (and their multimedia producer colleagues) to ensure that they make the most of the possibilities of simulation systems.

Adaptivity. As a subset of interactivity, the ability of a system to adapt to its users activities is of great interest in assessment. We know from our work with adults and children in formative learning contexts in the UK, that a critical factor in engagement and interest is the perceived difficulty level of the task (along with perceived relevance, media richness, feedback, etc.) In science it is of particular interest in allowing the assessment task to be revealed gradually and in ways which depend on the learner's knowledge and preferences. From an assessor's perspective, an adaptive test allows more critical judgements about ability and knowledge to be made in a given time because the candidate spends little time on tasks that are too easy or too hard.

7 System Diagram

The diagram below shows the web-based distributed functionality of the current system.

High Level Item Bank Systems Diagram



5 Phase 2 – Developing the Concept

The item bank service is now operational in the UK (www.itembank.org.uk) and is being considered as an underpinning technology service for a variety of initiatives as follows:

Providing the primary revision and support service for Basic and Key Skills.

Currently revision support CDs are released on a regular basis to provide a range of new assessments and to ensure that candidates are revising using current test content. This is a costly and uncertain process as we know that many centres will not update their software regularly even if updates are sent out. The webservice (which provides downloaded items for offline and low-bandwidth use) is seen as the optimal way to provide immediate access to contemporary information.

Developing internal processes for producing and managing items.

The bank offers many possibilities for streamlining the production process which is currently a separate and substantial programme managed entirely on paper. By collaborating in the production of items designed from the outset for on-screen and paper usage, the service will save money in the item production process.

Providing screening, diagnostic and formative assessment tools.

There are many areas in which the bank can potentially assist with teaching and learning, building on its initial purpose and its current usage. This work combines educational innovation (for example in developing interactive learning plans, links to learning materials, etc.) with technical developments (such as adaptivity, innovative interactive items, rich media content, and personalised user interfaces), and involves trialling of results at every stage with learners. We have identified a range of discrete projects in this category that include:

- Adapting the items and tests to fully meet the needs of disadvantaged candidates (this is significant as many interactive and media rich items cannot easily be accessed in unmodified form by learners with disabilities).
- Adding detailed "Item Feedback" to all present items and all new items to extend formative use.
- Exploring validity and reliability issues in the use of adaptive algorithms based on sub-banks of content for initial assessment, screening and formative assessment purposes (linked to dynamic learning plans).

Evaluating the potential for connecting the Item Bank to awarding body processes.

Currently, items are developed by a central government service in batteries of pre-configured tests written as complete tests. These are then distributed to the 20+ awarding bodies (in pdf format) for use in their various examination programmes. The rapid move toward much more frequent test sessions (and, in the immediate future, on-demand examinations) places great pressure on production processes in terms of volume of items required, costs of production and speed of response. The project can potentially deliver pre-packaged statistically and contextually-balanced tests in standards-based formats for use by awarding bodies in delivering on-demand tests, while allowing test writers to move toward writing single items for bank refreshment.

6 Quality Assurance Issues

The quality assurance of high stakes public examinations in the UK examination system is, in most cases, based on a combination of:

- Exam paper development undertaken by expert developers
- An extensive editing and review process that ensures the paper is:
    - Valid
    - Reliable
    - Fit for purpose
    - Of a similar standard to previous papers and the qualification level generally so as to maintain standards over time
    - Balanced
- A process that ensures confidentiality and security of papers
- A post examination moderation and standardisation process that ensures that the grades awarded are reliable and consistent with maintaining standards over time.

The system assumes that, in any one examination series, only one paper is sat by all the candidates and that the paper is sat on the same date and at the same time by everybody. This enables the grade awarding process to be based on statistics and other evidence from the same paper and from candidates sitting the paper under broadly the same conditions.

This process has been followed for the Basic and Key Skills “live” tests until recently. On paper, the tests have been available on a monthly basis with all candidates taking the same papers, which have been moderated and standardised in the “traditional” way. However, the advent of on screen and on demand testing has meant that this process has had to be changed. For the present, the tests being delivered use material that has already been used in previous paper based tests and for which performance statistics are known

and pass marks available because they have been through the standardisation process.

However, the possibility of using an item bank for live tests, the use of innovative and other items that cannot have been pre-tested on paper and the extension of on demand testing will demand the development of new and, as yet, untried quality assurance processes for the UK examinations system to replace parts of the current process, particularly the post examination moderation and standardisation process. The use of items banks and on demand testing is not new. Many examples can be found in systems from other countries and a variety of techniques have been developed to ensure the delivery of tests that are balanced, have consistent performance characteristics and are valid and reliable. This paper is not designed to address all the possibilities, rather to identify the need to address the issue. The following techniques are not, therefore, either exhaustive or detailed but begin to highlight the possible strategies that will need to be adopted. These might include:

- Extensive pre-testing of items with sample cohorts. Unfortunately, because the UK does not have a tradition of pre-testing items, it becomes difficult to find centres willing and able to find the time and resources to undertake pre-testing and to ensure that the cohort is representative of the wide cohort taking live Basic and Key Skills tests.
- Using expert trials. Expert trialling, combined with the use of expert test developers provides one way of test development that strengthens the current expert development team approach.
- Seeding new items into live tests. A number of test systems use live tests to trial small numbers of additional new items, scattered through the test for which the performance statistics are collected but which do not count towards the candidate's own performance. A test of 40 items might, typically, have a further 5 items added.
- Developing more sophisticated algorithms. One of the key objectives of the item bank project is to test the performance of the algorithm in delivering balanced, valid and fit for purpose tests. Provided the item performance statistics are sound, the project expects to be able to fine tune the current algorithm to deliver such tests and to provide a model for further development.

As the use of item banks extends to different types of items and subject areas, further issues will need to be addressed. These will include the generation of paper for qualifications requiring a range of grades, not just pass or fail, items that might be used in tiered paper qualifications (e.g. GCSE where more than one paper is used to cover the whole grade range from A* to G), qualification where to paper uses a wide variety of contexts from one series to the next.

The use of simulations and adaptivity further complicates the process of quality assuring the test – consideration now has to be given for the varying ways in which the learner may progress through the test, whether that gives

them a valid opportunity to demonstrate their abilities and whether the outcome level or grade is valid, could be replicated by the candidate in another test and can be shown to maintain standards over time.

The UK regulator for England, the Qualifications and Curriculum Authority, has suggested that it would like to see the first on demand, e-assessment GCSE papers by 2008. This project provides one opportunity to address some of the issues that will need to be resolved before this objective can be achieved.

7 Conclusions

This project is still at an early stage of its potential development. Already, however, the potential for use in a wide range of formative and summative test assessment contexts can be seen. It is a good example of using a research and development strategy to deliver a service; in this case a wide range of formative testing material for Basic and Key Skills learners, whilst addressing all the development issues associated with the possible use of the system to deliver high stakes summative tests.

The Qualifications and Curriculum Authority (QCA) has set the English examinations and testing industry some tough objectives in asking it to deliver 75% of Basic and Key Skills tests from 2005 an increasing number of other e-assessment examinations from 2005 onwards and on demand testing from 2008. Similar objectives are being set by QCA's counterparts in Scotland, Northern Ireland and Wales. Item banking will form part of the solution to meet those objectives and the outcomes from this project will help address a number of the issues raised.