e-Assessment 'virtual realities' – the benefits and challenges

Patrick Craven

Cambridge Assessment

Abstract

A review of the benefits and challenges encountered during a live national deployment of an e-testing solution. This case study seeks to share with delegates the very real opportunities and barriers faced when developing and running an e-assessment service. Although the case study features a certain type of e-assessment and range of deployment environments it is hoped that there will be lessons and advice that many involved in e-assessment can relate to.

The exact details of the test and target audience shall remain anonymous but a broad outline of the nature of the assessment and profile of testing centres is provided to allow readers to benchmark against their own situation. The assessment took the form of a timed test with two item types. One item type was multiple response and the other required more sophisticated interaction with the ranking of a series of options, via a drag and drop interface.

The test environment displayed a timer and allowed for flexible navigation plus an indication to the test taker of whether items had been fully or partially attempted. With approximately sixty items in the test the display of the navigation features was given careful consideration to avoid unnecessary use of the screen 'real estate'. The screen 'real estate' is a phrase used to describe the available space on a screen when displaying components of a question – rubric, stem, options, navigation etc. In addition to this there was also the availability of a glossary (to aid with certain domain keywords, not part of the assessment focus) and Help function (to give users extra guidance on how the item interaction should work). The final feature of the

test was an initial instructions page where additional candidate details could be captured to provide extra information for research and analysis of the test performance. This page could be accessed in advance of the test but no further access to the test was possible until the start time for the test was reached.

The tests were delivered to eight educational institutions across the UK and through ten different locations. Several hundred candidates sat the tests but there were no more than 60 in any one location and no more than 150 across locations at concurrent times.

The tests were delivered through a 'secure browser' interface to ensure that no access to the Internet or local applications was available. The 'secure browser' and 'invigilator console' required some local installations to be made at each institution. The tests themselves were delivered online and so a resilient and robust internet connection was required for the duration of the test – 2 hours.

The test sessions used generic but sequential login IDs that were unique to each institution. These were administered by local invigilation staff and the candidate identity captured during completion of the initial information screen.

The 'invigilator console' was used to check all initial connections and monitor the status of each candidate. All candidates could be paused, restarted, stopped and submitted from this interface. The system has the capacity to allow extra time and a controlled range of text size and colour combinations, although none of these features were used for this pilot.

This paper reflects on the range of benefits and challenges that were encountered during the lead up to and deployment of the 4 week pilot. The case study is broken into three sections that broadly define the stages of the assessment process, these are: production, set up, session. In each area we consider three factors: people, technology and assessment. The final section offers some conclusions about how different assessment scenarios may be impacted by what was learned from this pilot.

Production

The introduction of choice around mode of deployment – paper or digital – is causing many in the educational and training industries to reconsider how they commission and produce their materials.

Traditional models of test production have been largely 'linear' in notion with a whole test being commissioned and produced until it is used as a 'one-off' instance in a timetabled exam. This model of production is wasteful and made even more so if common elements begin to follow two tracks – paper and digital. Where possible e-testing production systems should move to a more atomized mode of commission where content is created at a disaggregated level. Publishers should also seek to identify items and content that can be used in both paper and digital format and only state the final mode format when publishing the whole test.

Using these techniques we can begin to take the pressure off the resource that is used when creating content and also the associated deadlines for creating and publishing tests. The content resource becomes a 'well' that is constantly replenished through commission. Test construction becomes a separate but related activity drawing on the 'well' of material and using a different set of personnel, skills and expertise. The technique is more commonly known as 'item banking' and much has been written about it in the literature. This does require changes to the way that

the assessment material is commissioned, vetted and how that then feeds the test construction process.

Authoring

Creation of content for an e-test can include direct engagement with the test creation system, remotely or locally, and this can be very beneficial. In traditional models the usual procedure is for content to arrive on paper or more recently word processing file format.

By giving authors direct access to the production system much time and effort can be saved in the initial data entry and submission of draft content. Several days effort can be saved and the cost of postage is removed from the process.

It is also advisable to standardize as many elements of the production system as you can to ensure consistency and reduce the editorial burden later on. Doing this will also place less demand on your pool of authors, as it is easier to train them on the use of templates than all the features of a test authoring system. In our experience the transition to digital content creation has been less challenging than we may have imagined. Most content producers are well-versed in the use of email, applications and the Internet and so use of a content production interface was not a major issue. As a general observation it seemed that age of author had little to do with the rate of adoption but it had more to do with their inclination to change habits and their view of technology.

Another useful technique is to separate out those elements of a test or task which can be viewed as test independent. So this might be item rubrics, test instructions pages, data entry screens, results page etc. In this way they can be created as a library of assets that can be selected at test construction stage and modified if needed. The same approach is typically taken to production of images and audio and the guiding rule is that 're-use' is a key feature of e-assessment production techniques.

It should be noted that careful thought must be given to the range of attributes or tags that should be added to all digital content. This must be done at the design stage as it will inform many aspects of test construction, analysis and evaluation later in the process. The ability to include a range of directly related metadata against test items is a distinct advantage over paper tests, as it allows for much faster and more efficient construction and evaluation activity. This paper will not dwell on the range of attributes that can be assigned but for the purpose of the case study they fell into two broad categories – syllabus information and performance data.

In the case study mentioned here only two templates were required and this made production relatively easy. The template provided simple box areas for entry of stem and options, with additional spaces for answer key and scoring information. In this example the scoring information was added by internal staff but all other information provided by the author.

An initial level of quality assurance is applied to all the items/tasks, which ensures that only those of a certain standard make it into the 'well' for subsequent test selection. One very important feature of an e-testing system is the ability to preview an item, and later a test, as this forms an important aspect of the authoring and approval procedure. Consideration should also be given to the screen resolution the content is developed for and you must plan for a range of common resolution settings.

This is also the stage when initial accessibility features should be considered. Text sizes and any adaptations for colour changes etc. These features should reside within your e-test production system but the flag to check against them is a procedural matter. The preview of an item may standardize around a default brand to focus on general layout and item interaction but if you are then applying some additional branding the test will eventually need to be quality assured in that format.

In our experience this has added some quality assurance steps to the process. In traditional models it would be sufficient to sign off a printed version of a test that would serve as camera-ready-copy for a printer. It is reasonable to assume that what is passed across to them is what will be printed, although post-print spot checks are still required. In an e-test environment you need to check that items render as expected and then also that a whole test renders as expected. In our case we were using a controlled secure browser for delivery and so there was no need to view the test through a range of different browsers. If that is not the case you will need to plan for such a range of user acceptance testing.

These are clearly aspects that become more of an issue the wider and larger your scale of deployment. If the objective is to develop e-test material which is only to be accessed within your department or institution then the range of deployment environments becomes less of a challenge.

Set up

The main difference here is that new departments will become involved in the process and it is important that they are included early on in planning activities. Traditional models of deployment have usually involved assessment personnel (academics) and operational staff (dispatch and logistics etc). These two groups are still involved, although the operational role may diminish or change, but an important new role is that of technical support and this will involve IT/IS personnel.

This can be further complicated if the IT/IS function is partially or completely outsourced. This is not to say it is always complicated but it can make the lines of communication and agreement more protracted. In our experience this works best where the traditional users are given high levels of access to the IT systems and consequently the responsibility for final delivery. If too many people have primary responsibility it can lead to a fragmented deployment.

Production staff will need to allow for 'dummy' testing in a development environment and in the live environment. Over time as you become more confident with your systems this level of quality assurance might be reviewed but initially this will form an important additional set of tasks.

Whether you are delivering locally or externally the IT/IS department become a fundamental new contact point when preparing to run a test session. Again this is an aspect of the process that should be much easier if you are only coping with a local deployment – here one of the key factors is early engagement in planning activity for the whole deployment team.

Externally there are new challenges as the traditional point of contact would probably be an exams officer or equivalent management administration role. The traditional range of rooms used for conventional tests may no longer be appropriate. Access to a sufficient range of IT may not be possible and may create timetable clashes where that equipment is primarily used for lessons. These are not insignificant cultural and practical challenges that need to be addressed. The difficulty for an external supplier

is that you have relatively little influence over how such issues are resolved. In our experience we are starting to see greater recognition that IT equipment should be allocated to assessment and learning activities and all of the institutions we worked with on this case study had some prior experience of e-testing.

Advance communication and preparation becomes critical with e-assessment. It is vitally important to get access to the IT/IS staff as early as you can and if possible secure some direct channels of communication. This will speed up the process and help to ensure that the correct information is being transferred.

Delivering an online solution to external users is relatively simple so long as they can guarantee resilience of service – this is an area where we have seen a positive trend of improvement in recent years. The range of browsers is not too broad but you should take the time to understand what your target audience uses and build that into your UAT plans.

If you e-test solution requires some local installations time must be allowed for this to happen. In our experience you should allow as much time as you can for this activity. External institutions usually address IT/IS support in two ways – they will involve local personnel or it will be an external managed service. Although there are variations in between that is typically what we encountered, even with a range of institutions as small as eight. Establishing the initial contact took some time but once that was in place the rollout and deployment of the solutions was relatively simple. Another important factor is the degree of confidence you have that the local installation has been performed sufficiently well to cater for the pilot test session. This is not to say that the quality of local deployment is poor but highlights the inconsistent nature of IT systems and time available for local IT/IS personnel to devote to a pilot. This may have been different if it was preparation for a live test.

In many cases the initial test of installation and running a 'dummy' test did not take place in the location that was to be used for the pilot. Sometimes the local IT/IS staff did not know at that stage what the exact location would be, sometimes they did not have access to it at that point in time. This did lead to a couple of issues as test dates approached and local IT/IS staff encountered problems.

Despite all of the planning and preparation there may still be issues on the day and over the period of a relatively small pilot these are some of the issues encountered:

- one room/location lost all internet connection on the day but paper-based contingency had already been planned for. The e-test system is capable of local offline deployment as well but as that would involve some setup on the day the decision was taken to use paper as the backup option. Issues of this scale, as with powercuts, are very difficult to recover from.

- several instances of hardware failures – keyboards, mice and monitors – test preparation guidance states that additional devices should be available and this catered for all these instances.

- two instances of OS failure at boot-up that would have taken too long to resolve on the day. Spare capacity PCs coped with these issues.

In all cases the actual deployment of the solution was relatively quick and straight-forward despite initial reservations on the part of many of the IT/IS personnel. Where internet connections are required some clarity should be sought over whether the institution has 'approved site lists' for URLs in operation, whether they have any anti-virus software that may interfere with activity, if the ports selected for use are available and whether the level of access is sufficient to run the tests at the scale

expected. Scale and access issues should be considered alongside any other activity taking place in the institution at the time of the test.

The supply of a test can be controlled to a very fine degree and can be made as secure or as open as you want it to be. Although the supply of test material can be left until the last minute we allowed access to the encrypted test material 48 hours in advance of the test to allow for scheduling and registration of candidates. A major benefit of an etest system is that access can be granted in stages – with a traditional system once the paper is dispatched all that protects it is a sealed envelope. The final trigger for access to an etest is a combination of the candidate login and the timetabled start time and so the whole system becomes far more secure than paper.

Changes to content can also be made within minutes and there was one instance of a change to wording in an option requested less than 24 hours before the first test session. This was done and deployed in minutes but in traditional terms would have resulted in the production and issue of errata or even the disqualification of that item from the marking process.

Session

In a traditional examination model the exams officer will usually hand over to the invigilator to run the test session. In the past these have been academics or teachers but increasingly they are external, contracted staff brought in just to administer test sessions. This entails checking in candidates as they arrive, reading out instructions, monitoring activity during the exam and collecting in materials at the end of the session.

Here the invigilator is presented with a new challenge – to run and monitor a set of e-tests using a console screen. A common misunderstanding is that this means that they no longer walk the exam room but this is not the intention. The console is used as an additional device to check progress and connections plus control any pauses etc. For the pilots there was limited opportunity to train or brief invigilators on use of the system and so we had personnel at every session to observe and offer assistance if needed. After an initial overview it was pleasing to see that little additional support was required and all were relatively comfortable with the operation of the console. Many of the questions that were raised were about etesting in general, such as security, malpractice and collusion.

The inclusion of IT/IS staff to support the exam session is a new requirement. For the pilots they were typically available at the start of the day during the initial setup and test launch period and thereafter were 'on call' throughout the duration of the test. Some decided to stay throughout the test session to observe what happened but in most cases no further assistance was needed. All of the issues we encountered during this case study arose before the start of a test (hardware and software etc) and the most serious was the loss of Internet connection. This impacted the whole site and so was never going to be resolved in time to resume the etest.

Institutions noted that the delivery and support of e-assessment (and specifically timetabled tests) would require much closer collaboration between administration, academic and IT/IS staff than was needed for traditional exam sessions. This was not felt to be a problem but may require a review of internal procedures and training programmes. It is worth noting that this relationship is again easier to manage if the e-assessment solution is only to be delivered internally. For external solutions the

supplier is very reliant on adequate resource being made available and to ensure that they are suitably briefed.

No major issues were observed with candidates' use of the etest interface and that was with relatively little or no time to engage with practice tests. The age profile of the candidates was late teens to early twenties and perhaps this level of comfort with digital interfaces is to be expected – it was reassuring to have the assumption confirmed. The majority of the candidates commented on a preference for the on screen form of assessment but also some expressed no strong preference simply seeing the test as a necessary, regardless of mode.

At the end of the test the invigilator could observe that all tests had been closed and received confirmation of submission. We were able to check almost instantly for receipt of the files, all of which provides a significant benefit over existing postal procedures. If the test form had of allowed for automatic marking and the items had been pre-calibrated there is no reason why results could not have been issued instantly.

Conclusions

In this example case study there is no significant difference in the type of question presented to candidates. Multiple response is a common item type and rank order items are used in paper-based tests, although the form of interaction is different.

The benefits and challenges of the etest in this case study mostly relate to areas of back office production and delivery but these are important considerations for any organization looking to deploy e-assessment solutions.