

Easing the transition from paper to screen: an evaluatory framework for CAA migration

Mhairi McAlpine*

Scottish Qualifications Authority, Glasgow, UK

Computer assisted assessment is becoming more and more common through further and higher education. There is some debate about how easy it will be to migrate current assessment practice to a computer enhanced format and how items which are currently re-used for formative purposes may be adapted to be presented online. This paper proposes an evaluatory framework to assess and enhance the practicability of large-scale CAA migration for existing items and assessments. The framework can also be used as a tool for exposing compromises between delivery mechanism and validity—exposing the limits of validity of modified paper based assessments and highlighting the crucial areas for transformative assessments.

Background

All holders of assessment materials are currently investigating what the impact of ICT will be on their future practice. There is an acknowledgement that the traditional manner of assessment will change, however as yet no clear vision of what will replace it. Bennett (2001) outlined the major changes that assessment would undergo in response to the changing technology. Ripley (2003) has built on this idea to refine the three models of change which will dominate in the migration from paper-based to screen-based assessment. Figure 1 gives an illustrated summary of the Ripley model.

There is, however, an issue in how we get from a mass system of testing to an individualized assessment structure, quite apart from the changes that the introduction of ICT will create. While the introduction of ICT is, for many, a time to rip up the rule book and start again—there is a need to take the practitioners along with the technology, starting from where we are now and introducing change slowly and incrementally

*Scottish Qualifications Authority, Glasgow, UK. Email: Mhairi.McAlpine@sqa.org.uk

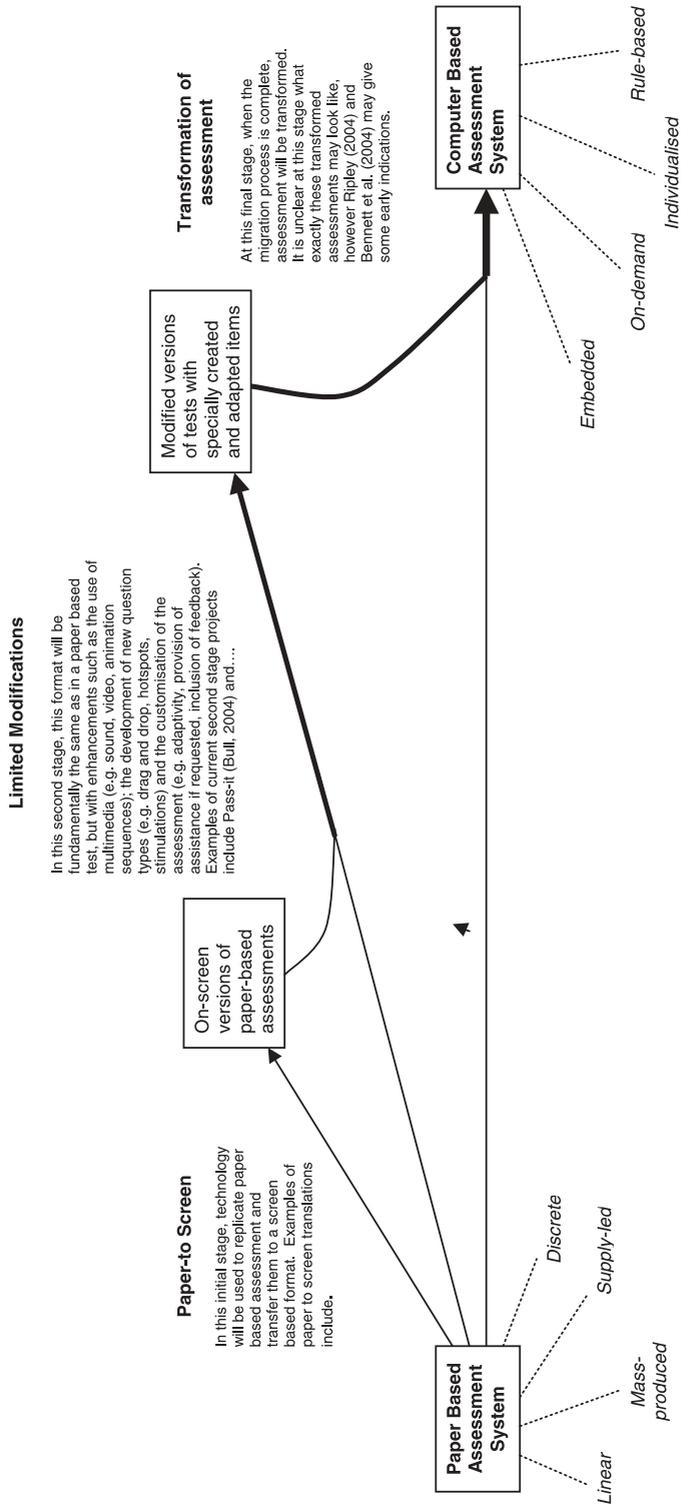


Figure 1. The Ripley model

and ensuring that the cultural shift between paper-based and computer-assisted assessment is supported—and above all that the confidence in assessment systems is retained.

This paper looks at how existing systems of assessment and collections of items can be put online and how these can be evaluated to appreciate the difficulties and challenges that this transition presents. By acknowledging where compromises have been made and differences created, there is an awareness of the limitations of the technology. This can expose where validity risks are being compromised for the sake of the assessment format. This proposed framework also can be used as a means to prioritize developments and make explicit where the likely challenges in certain types of developments may be. Used in this manner, it can be an important tool for implementers of large-scale CAA developments to manage change from paper-based to on-screen delivery.

Most UK developments are currently in the first or second stages of the transition process, looking at how they can adapt their current assessment practice to an on-screen delivery format. Although this might seem a little un-ambitious, especially when compared with the more radical online assessment methodologies being developed, it is a necessary evolutionary step to engender confidence in computer delivery without too much of a radical change in the assessment format.

The Scottish Qualifications Authority is currently in the process of exploring the potential of CAA (McAlpine & Ware, 2003). It is anxious to avoid the fragmented approach that McKenna and Bull (2000) report has characterized the development of CAA in higher education in the UK and the resulting difficulty in achieving sustained systematic innovation across the education system. To that end it has actively sought partnership in its CAA activities with its stakeholders, and is involved in setting up the infrastructure which underpins CAA systems, and putting in place the processes of change which will ease the transition for all involved. We work in partnership with our centres and the rest of the Scottish Educational community and are keen that we are aware of what we are expecting of them through this time of innovation, and are doing all that we can to support them as they make this transition with us.

Introduction

Six subjects were chosen to form the medium of this pilot. These were English, Maths, a science (Chemistry), a humanity (History), an art (Music) and a modern foreign language (French). All were chosen from the external assessment component of the Higher. These are summative terminal assessments which are typically done by more able students in their fifth year of High School in what is considered the main determinant of entry to Scottish Universities.

The question papers varied in a number of respects: the time allocated to the examination; the number of questions; the response that the questions demanded; the appearance of items, and the supporting material associated with the papers. It was considered that the major change issues with a move from paper based to on-screen

assessment would be the response type that an item in the paper expected, and the inclusion of any stimulus material which is currently given on paper.

Each item in each of the question papers was considered—looking at any stimulus associated with the items, the type of response input that the item required, and the type of marking required by the item.

Classification of ease of migration

In order to facilitate decisions of which types of items should be considered suitable for straightforward translation, which for modification and which should be considered from a transformative standpoint, A coding was established to identify which types of responses; input mechanisms and stimulus material were able to be directly translated into an computer based format. A classification scheme was developed to identify the extent of the challenge (Table 1).

As can be seen from Table 1, classes one and two are most suitable for direct on-screen migration, classes three and four were possible to implement with some consideration, but may well be more suited to a modified form, while classes five and six were not available for direct translation and may require the kind of third stage transformative work.

Table 1. Classification of ease of migration

1 Currently widely available

There should be only trivial issues to resolve. Immediate implementation is feasible.

2 Currently available, but requires refinement

Some minor decisions may have to be taken about how exactly it is implemented. Immediate implementation is possible, however small amounts of work, or consideration of issues may have to be given to ensure long-term success.

3 Currently available, but needs development for operational use

Substantial decisions may have to be made about the technology used or the manner in which it is implemented. It may require investment to ensure that it is of the standard which we would require.

4 Limited availability and requires development

Decisions would have to be made about how it is developed and to what extent. Implementation will not be possible until the development is complete. It will require some investment for operational use.

5 Potential availability with commitment to development

This technology may well be at the beta stage or only available as a trial version. Substantial decision would have to be made about how exactly it is implemented and in what form, it will require investment both to finalize the technology and to make it operationally available.

6 Not currently available without significant commitment and investment

There is no reliable method of doing this at the present time. Experimental projects are at an early stage or have not reached satisfactory conclusions. In order to implement this operationally, significant resources would have to be deployed to ensure its success.

Table 2. Count of stimulus by code

| Stimulus code | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|--------------------------|------|------|-----|------|------|-----|-------|
| Count ^a | 131 | 41 | 6 | 84 | 31 | 8 | 300 |
| % of items | 43.7 | 13.7 | 2.0 | 28.0 | 10.3 | 2.7 | 100 |
| % of items with stimulus | - | 24.2 | 3.6 | 49.7 | 18.3 | 4.7 | 100 |

^aNote that one question had both a Photo and Quote stimulus.

Description and classification of items by attribute

Stimulus

From the papers selected, five types of stimulus material were identified: diagram/graphs; photo/drawing; quote; aural cue and formula. Each question was classified according to which type of stimulus was associated with it, the most numerous minority (43.7%) of the items had no stimulus material associated with it, while only one question had more than one associated stimulus (Table 2).

Each stimulus code was taken in turn to consider how difficult it would be to migrate that type of stimulus to a computer format. Table 3 details this classification together with some analysis of how it was reached. Issues which were considered included how difficult it would be to present through computer, how difficult it would be to access it, how candidates with special needs might be affected by this method of delivery, any special pieces of software which might be required to enable this, what the 'industry' tended to use for the delivery of this type of material, and alternative ways that it might be presented, including some evaluation of these methods.¹ To construct this classification, a number of approaches were used, including a literature review, consideration of software known to the author and consultations with those involved with practical projects involving CAA. It does not claim however to be a definitive account of all available technologies.

Response type

From the papers selected, six types of response type were identified: numeric responses; algebraic responses; lexical responses; diagrammatic responses; closed responses and selected responses. Each question was classified according to which type of response was expected from it. The most numerous grouping of responses were lexical responses (class 3), accounting for 59.6% of the responses required. Categories two and three were further divided by the length of response expected, as in these response types that is a significant factor influencing the ability of the computer to automatically mark the item (Table 4).

As with stimulus, each response type was taken in turn to consider how difficult it would be to migrate that type to a computer format. Two issues were considered, how difficult the response type was for a candidate to enter an answer into

Table 3. Description and classification of stimulus types

| Code and description | Stimulus classification |
|--|--|
| <p>Code 1: Diagram/Graph</p> <p>Items with images where the information in the stimulus was essential to the answering of the question—thus an accurate reproduction of the image would need to be rendered on computer in order not to disadvantage candidates.</p> | <p>Class 2</p> <p>There are a number of standard image formats available that diagrams and graphs can be rendered in. Most CAA engines and VLEs will accept these forms, however the display mechanism may cause subtle variations, which may affect question quality.</p> <p>Furthermore the capabilities of the machines which the candidates attempt the question may affect the rendering.</p> |
| <p>Code 2: Photo/Drawing</p> <p>Items with images where the information in the stimulus was impressionistic on the answering of the question. Thus so long as the image was visible and retained its meaning it would be an acceptable rendering.</p> | <p>Class 1</p> <p>There are a number of standard image formats available that photos and drawings can be rendered in. Most CAA engines and VLEs will accept these forms.</p> <p>Again the capabilities of the machines which the candidates attempt the question on may affect the rendering—and although this might not be so critical as in the above example, it may bias results towards better resourced centres and candidates.</p> |
| <p>Code 3: Quote</p> <p>Items with text stimulus of a few words to a sentence or two. As with the above this could be rendered as an image, however it is assumed that the underlying coding style is textual.</p> | <p>Class 1</p> <p>Most CAA engines and VLEs will accept textual stimulus material.</p> |
| <p>Code 4: Aural Cue</p> <p>Items which required candidates to listen to something before responding.</p> | <p>Class 3</p> <p>There are a number of rendering mechanisms for aural stimulus material, however the implementation of this into live examinations is surrounded by technical and practical issues which would have to be resolved prior to live use.</p> <p>Some of these issues include</p> <ul style="list-style-type: none"> ● the sound quality which, as with the image rendering, may be affected by the specification of the machine on which the candidate is being examined; ● the candidates control over the music playing – can they play it themselves, or would the computer play it for them in the manner that the invigilator currently does; |

Table 3. *Continued*

| Code and description | Stimulus classification |
|---|---|
| <p>Code 8: Formula</p> <p>This code was used to demarcate stimulus which was presented in a standard subject specific form (in this case using mathematical notation and chemical notation). These could, in theory at least, be presented as an image and fall into code 1, however essential information would be lost which should be retained to maximize the usage of the question (not least in question generation).</p> | <ul style="list-style-type: none"> ● how candidates might have access to the aural cues at different times without disturbing other candidates in the room; ● may a computer be able to get round some of the problems that candidates with SEN may have in accessing certain part of the examination? (e.g. through increased amplification etc.). <p>Class 5</p> <p>There are a number of ways that chemical and mathematical formulae can be represented on computer. These include LaTeX; MathML and ChemML as well as a variety of plug-ins. None of these, (except LaTeX, which is an imperfect partial solution), can be adequately rendered on the majority of CAA engines or VLEs—this would cause a significant problem should the meaning behind the formulae have to be retained.</p> |

the computer, and how difficult it was to enable automatic marking of that type of question. Table 4 details these classifications together with some analysis of how they were reached. Issues which were considered for input purposes included special characters, free-input, specialist notation and some consideration of accessibility issues. Issues which were considered for marking purposes included the availability of technology that could enable computer-based marking of these types of questions, any minor changes that could be made to the questions to make them easier to mark on computer, and the reliability of the marking. As with stimulus, it does not claim to be a definitive account of all available technologies, but based on existing practice, known issues, currently available software and published literature.

Visualizing the papers

Using the stimulus class and the highest of the marking class and input class, a view of how easy each of the papers would be to migrate to CAA were established. A short consultation exercise in which method of visualisation could most easily communicate the essential information indicated that bubble graphs were favoured over the two alternative methods put forward (stacked area graphs and luminosity squares).

Table 4. Description and classification of response types

| Code and descriptions | Input classification | Marking classification |
|---|--|---|
| <p>Code 1: Numeric</p> <p>Items where a numerical answer was required.</p> | <p>Class 1</p> <p>In the case of most numbers, input is fairly straightforward and can be done using standard notation on a standard keyboard—additional characters which would be required in addition to digits would be ‘-’ (negative numbers) ‘/’ (fractions), ‘.’ (decimals), ‘i’ (imaginary and complex numbers), ‘e’ (2.11) and ‘π’ (3.14). The only one which causes significant issues and is not found on standard keyboard is ‘π’. Most CAA engines accept numeric data, input issues should be minimal.</p> | <p>Class 2</p> <p>There are a variety of ways that numerical questions may be marked. Sometimes a precise answer is required</p> <p><i>(e.g. what is $1/2$ expressed as a decimal?</i> <i>ANS = 0.5 only)</i></p> <p>and other times more than one representation may be acceptable</p> <p><i>(e.g. If 8 cakes are shared among 10 people how many do each get?</i> <i>ANS = 0.8 or $4/5$ or $8/10$)</i></p> <p>Thus there would have to be the facilities available for the evaluation of the answer and an understanding of numerically equivalent forms as well as the facility to limit the acceptance of equivalent forms in certain cases.</p> <p>There are a number of CAA systems which have both of these capabilities and although tweaking them to the precise requirements may require some work, this should not be a significant limiting factor.</p> |
| <p>Code 2: Algebraic</p> <p>Items where an algebraic answer (i.e. one including unknowns, typically represented by letters) was required.</p> <p>For marking purposes it is separated into sections 2a which requires only one line of input (the answer expected</p> | <p>Class 3</p> <p>Where unknowns are represented by letters as is common—this should not pose a problem as they are found on a standard keyboard. At lower levels unknowns may be represented in other forms (e.g. stars or question marks) which may prove more challenging.</p> <p>Algebraic answers can quickly become complex and may require the whole range of algebraic notation available. This might include (but not limited to) complicated fractions, integrals, logs and powers. There are a number of ways of</p> | <p>2a—Simple Algebraic: Class 3</p> <p>As with numeric answers, algebraic answers may have a number of acceptable equivalent representations however the question may limit the number of acceptable forms. Accepting too many equivalents may compromise the question, especially where the question is designed to test their ability to manipulate algebraic expressions. Thus, as with the numeric questions, there would have to be the facilities available for evaluation of the answer and an understanding of numerically equivalent forms as well as the facility</p> |

Table 4. *Continued*

| Code and descriptions | Input classification | Marking classification |
|--|---|---|
| <p>would be a formula) and 2b those which would require more than one line of input (the expected answer would be a proof).</p> | <p>inputting these as well as standard text input, such as selection from a menu (as in MS Word Equation Editor); special codes (as in LaTeX); or typing in characters in an appropriate order. Many CAA engines are able to accept algebraic data, however the quality of their input mechanisms vary (especially for more complex expressions) and input issues may challenge validity if they are not adequately considered.</p> | <p>to limit the acceptance of equivalent forms in certain cases.</p> <p>Where input issues were affecting the quality of the answers, there would have to be recognition of common errors which had been caused by input difficulties. This would best be recognized at the input stage – thus it might require an intermediate evaluation of the answer given checking for common input errors (e.g. x^2 computer asks if that is x^2; $2x$ or ‘times 2’).</p> |
| | | <p>Class 2b—Proof: Class 4</p> <p>There are a variety of packages available which allow for input of algebraic expressions longer than one line, in some cases however this would have to be coded as separate answers. The CUE system in use with the Pass-IT trials, allows supporting ‘steps’ to be accessed and used when candidates request them (this would be of particular relevance in the case of codes 1a and 2a where the answer itself is of a different form, however the proof may allow access to the partial credit available)—although it could be effectively insisted upon.</p> <p>There would be a number of difficulties associated with proofs, particularly as there may well be no one correct proof, but a variety of answers which may legitimately gain the available marks. McCabe (2001) has suggested the objectification of proof questions to assess this type of learning and suggests a variety of ways which various CAA engines have approached this. All in all, this would be a problematic area and one which would require further consideration.</p> |

Table 4. *Continued*

| Code and descriptions | Input classification | Marking classification |
|---|---|--|
| <p>Code 3 Textual Response</p> <p>Items where the response expected was textual. This may also include non-lexical answers (such as a telephone number or date).</p> <p>The code is divided up into five subsections dependant on the length of response expected for analysing the potential for computer enhanced marking.</p> <p>Code 2a examines single word responses (which may on occasions cover a short phrase or non-lexical responses such as a date), code 2b covers short responses, ranging from a few words to a sentence; code 2c covers expected responses ranging from a sentence to a paragraph. Code 2d, codes extended responses where between one and three would be expected and code 3e covers</p> | <p>Class 1</p> <p>The input mechanisms for this response type would be fairly straightforward—involving the standard characters on the keyboard—although this may include numbers and other characters (such as ‘£’; ‘/’ etc.). This would be accepted by all CAA engines—although any unusual characters which were likely to be used in an assessment would have to be flagged and considered.</p> <p>As the length of response demanded grows it would have to be considered whether there were sufficient elements in place in the case of a systems failure and whether there could be checks built into the system to ensure that no input was lost.</p> | <p>Code 2a: Single Word: Class 2</p> <p>Almost all standard CAA packages mark single word responses. Problems may well occur, especially with less able candidates where spelling is poor, compromising the computer’s ability to recognize the answer, or where there are a number of synonyms which would be equally acceptable. These can be circumvented by entering a variety of alternative answers, including common spelling mistakes, which should also be marked as correct. Alternatively, for spelling errors, a formulaic interpretation can catch unusual errors, however this must be monitored carefully to ensure that the net is not being cast too wide.</p> <p>Changing the format of the question may be an option worth consideration in some cases—there are questions would lend themselves to objectification (perhaps through pull-down menus or drag and drop). Real-time spell checks may also assist candidates enter a response which was recognisable to the computer, with mis-spelt words being highlighted to the candidate for revision – suggesting alternatives may however compromise the validity of the test.</p> <p>Ultimately it should be fairly straightforward to computer mark single word response, however there may need to be an element of human marking back up, or question re-design to ensure the reliability of the system. These types of items could probably be migrated with minimal difficulty.</p> <p>Code 3b: Short Response: Class 3</p> <p>These responses tend to be relatively factually based—where the mark key</p> |

Table 4. Continued

| <i>Code and descriptions</i> | Input classification | Marking classification |
|--|----------------------|--|
| <p>essay responses where a response over two paragraphs would be demanded.</p> | | <p>is determined very much by the content of the response, rather than by its construction or style. These should thus not present too much of a challenge to most CAA systems although the methods of obtaining a reliable marking schema which can be entirely computer driven may be laborious and time consuming. For small entry subjects, the process of creating the algorithms for computer marking may negate the benefits of on-line marking unless progress is made in this area. The technology to make this possible is certainly available, however some advances would help to make it a desirable innovation.</p> |
| | | <p>Code 3c—Short Answer: Class 3</p> <p>Most CAA packages accept short answer responses, however the accuracy of the marking varies in its reliability. <i>Michell et al. (2003)</i> have suggested that there are systems available which after human moderation can mark at 99.4% accuracy overall. In a trial all items were marked at over 93% accuracy and with 98.1% of items over 95% accuracy. For more problematic items these could be redesigned to ease marking. They would still require a level of human moderation (figs above are post moderation) but this is a hopeful development. As with the above there is some issue with the time which may be needed to create and moderate the marking scheme, however as these question-types are more difficult to mark by human markers, this may not be such an issue in this instance. There are a few issues still to be resolved with these items however it looks as if reliable marking of short response questions may appear soon.</p> |

Table 4. *Continued*

| <i>Code and descriptions</i> | Input classification | Marking classification |
|---|---|---|
| <p>Code 9: Diagram</p> <p>Items which require the candidate to draw something which is then evaluated.</p> | <p>Class 5</p> <p>This would be a difficult one to implement without significantly changing the question or providing very specialist hardware, although questions did vary in their input computerisation difficulty. Much though would have to be given to what the question was actually designed to assess, and how this could be put onto computer without compromising the nature of the assessment. Some methods of input are already available—such as the</p> | <p>Code 3d—Extended Response: Class 4</p> <p>This type of response would be best marked in a manner similar to that described below - and would have similar problems and challenges associated with it—although it might be imagined that the problems associated with essays would be reduced as the size of the material was reduced, this may not be the case and further investigation into the technologies available would have to be performed.</p> <p>Code 3e: Essay—Class 4</p> <p>There are packages on the market which are designed to automatically mark essay responses. The most widely known and used is the e-rater system from Education Testing Services (ETS). These have problems associated with them, and it is not clear whether they would be accepted by markers and teachers. The developments in this area tend to come from the US, and are heavily influenced by US assessment practices, which may create challenges when migrating the technology to a Scottish context.</p> |
| | | <p>Class 5</p> <p>Much of the possibility of CAM would be determined by the input mechanism used. Where specific tools were used, there might be less difficulty in establishing a marking mechanism, however where a more generic input mechanism was used this may prove more challenging.</p> <p>There are already CAA packages which do allow some element of automatic marking of questions requiring a diagrammatic response, however most significantly change the</p> |

Table 4. *Continued*

| Code and descriptions | Input classification | Marking classification |
|---|--|---|
| | | demand of the questions in doing so. The pass-IT trials included one such questions, however it is unclear whether the question was indeed equivalent to the paper based form or an alternative way of assessing the same skills. A full review and evaluation of the area would have to be undertaken. |
| Code 10: Cloze Response | Class 2 | Class 1 |
| Items where there is a body of material with missing pieces together with a choice of pieces to complete the material. Candidates are asked to insert these pieces at the appropriate points. | Although the fundamental computerisation of a cloze response is quite unproblematic, there are a number of minor issues. The format of answering may be changed on a computer—including say drag and drop or scrolling and there may also be other methods of input. This may indeed add to the questions reliability by slightly altering the input mechanism to get rid of externalities such as spelling ability. | Although there will be trivial issues around spelling (where candidates are expected to type their response in)—these marking issues suffer from the same kind of difficulties as single word response items (code 3a). Most other marking issues will be similar to those with paper based clozes. |
| Code 11: Selected Response | Class 1 | Class 1 |
| Items which required candidates to make a choice out of a number of possible given answers. | Input for these types of questions should not pose any particular problem and could be implemented in a number of ways. Indeed computerisation of selected response items can add to the validity in a number of cases by changing the input mechanism (e.g. to a hotspot) rather than the traditional A/B/C/D response. | Techniques for marking selected response items on computer are well established. This should pose no particular problems. There may be issues where these items migrate from there traditional form to newer less well tested input mechanisms (e.g. hotspots) however these can be avoided until confidence in their reliability can be ensured. |

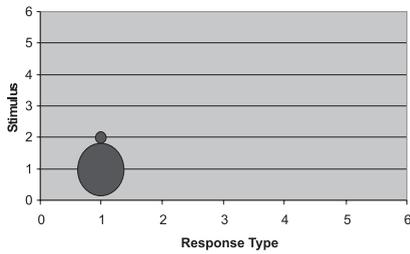
Table 5. Count of response type by code

| Response type | 1 | 2a | 2b | 3a | 3b | 3d | 3e | 3f | 4 | 5 | 6 | Total |
|--------------------|-----|-----|------|-----|------|-----|------|------|-----|-----|------|-------|
| Count ^a | 8 | 3 | 36 | 11 | 45 | 25 | 38 | 71 | 13 | 3 | 48 | 300 |
| % of items | 2.7 | 1.0 | 12.0 | 3.7 | 15.0 | 8.3 | 12.7 | 23.7 | 4.3 | 1.0 | 16.0 | 100 |

^a Note that one question was classified as both 1 and 4, one as both 3 and 5 and one as both 5 and 9.

1. Papers suitable for direct migration

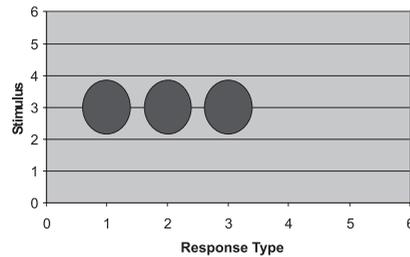
Chemistry Paper 1



Only chemistry paper 1 was really suitable for direct migration. This was a multiple choice paper and, as can be seen from the graph above, very few of the questions pose any difficulty at all in migration to a computer based format. This examination could migrate practically instantly with very little modification.

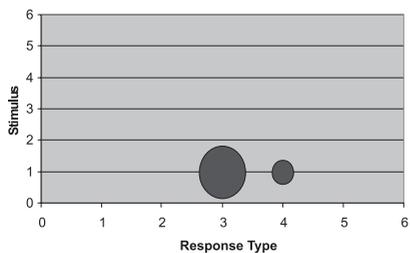
2. Papers which may be suitable for modification

Music Paper 2



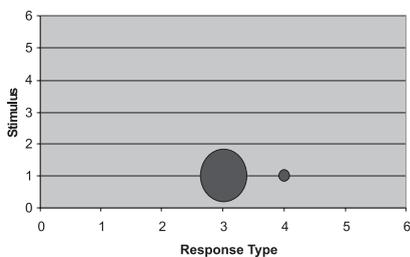
This paper has some challenges surrounding the stimulus material that it uses and indeed the stimulus is the biggest problem. The response types used for the majority of questions do not pose significant issues and the difficulties inherent in some of them could be circumvented.

English Paper 1



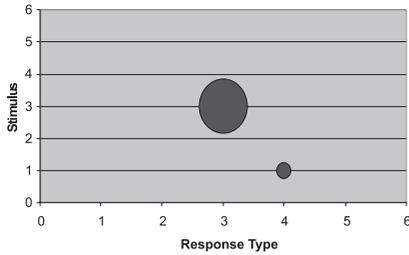
In this English paper, the response type is causing difficulties in migration although there are no stimulus issues. The questions are all fairly similar, suggesting that technical developments are needed to overcome the difficulties faced. Changing the response type may challenge validity unless carefully studied.

French Paper 1



This paper has some technical challenges associated with the response type used, although these are not insurmountable. Given the uniformity of the response types, hastening migration through adaptation may pose challenges to the validity of the assessment.

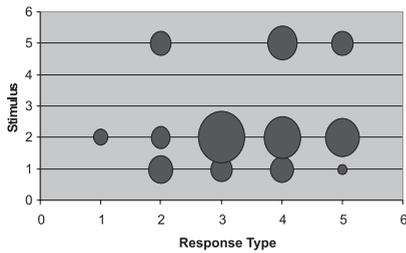
French Paper 2



There are some issues to be overcome in this paper both through the stimulus used and the response types used, although again these are not insurmountable. Whether these can be done whilst maintaining the quality of the assessment needs to be viewed in a wider context.

3. Papers where some of the paper may have to be rethought for CAA

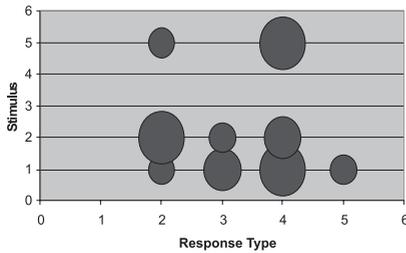
Chemistry Paper 2



The variety of difficulty levels and technical difficulties which need to be overcome, mean that this might not be a suitable paper for migration until a significant number of the technical challenges have been overcome. Should early migration be seen as desirable, compromises and adaptations to the response types used may be necessary. Migration of

some of the items may challenge validity.

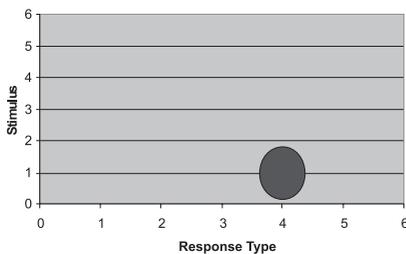
Maths Paper 1



This paper poses significant challenges to migrate. There may be some room to adapt the response types in particular and certain parts of stimulus could be presented in a more migratable format, however there are a number of items which would have to change significantly in order to be computerized.

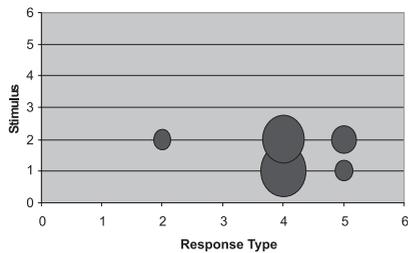
4. Papers that may require a transformative approach

English Paper 2



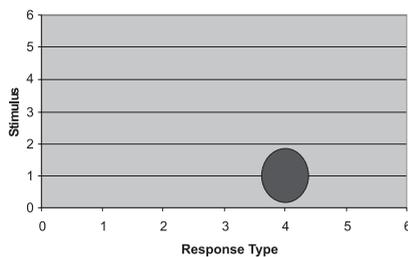
As with the above, the response type used in this paper is not conducive to migration. Technical developments are needed to enable migration. This may be an area in which the efforts involved in paper to screen migration may not be worth the transitional results.

Maths Paper 2



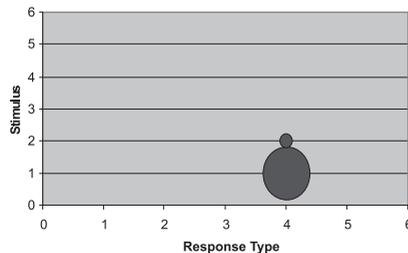
This paper also has significant challenges to migration. There are few issues with the stimulus required, however the response types used are not conducive to migration. The response types used would have to be reconsidered should migration be desired in the short term.

History Paper 1



As with the English papers this history paper has significant problems associated with the response types used. Using more migrateable response types would significantly alter the character of the exam.

History Paper 2



As with paper 1, considerable technical developments are needed to enable these types of items, and adaptation and changing response types would significantly alter the character of the exam.

Conclusions

This methodology suggests a mechanism whereby papers and subjects which are being considered for migration to CAA can be compared in their suitability for online delivery taking into account the wide variations in response types and stimuli which are found across papers and subjects. Furthermore it also gives an indication to what extent the migration to computer-assisted formats for assessment may pose challenges to both reliability and validity, and at the same time open up opportunities for re-thinking the methods of assessment in those areas. The subjects which cannot easily be migrated to a CAA format are perhaps the ones most amenable to the type of transformative assessment talked about by Ripley and Bennett, and the most promising for emerging technologies – however it must be considered to what extent this

is indicative of SQA's assessment practices and how can be allocated to the curricular areas themselves.

While the conclusions that can be drawn from only six subjects, at one level are limited, the classification framework is now in place to rapidly construct comparable indicators for other subjects and other levels. This will give us an indication of what issues need to be tackled in each subject, how significant a problem they are and (for institutions which hold large numbers of paper-based items which they would wish to move to an online format) suggest an order in which migration can commence.

One of the weaknesses of this project was that there was insufficient review of the emerging item marking and display technologies. It was outside the scope of this study to perform the type of comprehensive review which would be required and the classifications are given to the items should be interpreted with that caveat. A comprehensive study of emerging CAA technologies is long overdue and would greatly inform the sector, not least by ensuring that anyone using this methodology for migrating paper based items to a CAA format has a robust system in place by ensuring that the classification was as accurate as possible.

Notes

1. It should be noted that these codings are given on the basis of current knowledge of the author and are not based on any external categorisation. This may lead to inaccuracies in classification where technology has progressed beyond the author's awareness. Where an accurate classification is required it is recommended that a thorough review is undertaken and that further progress and development in the area is monitored.

References

- Ashton, Schofield & Woodger (2003) Piloting summative web assessment in secondary education, *7th International Computer Assisted Assessment Conference*, Loughborough, 8–9 July.
- Ripley (2003) The e-assessment tail: new opportunities? New challenges? The uses of ICT in assessment, *Beyond the Exam Conference*, Bristol, 19 November.
- Ripley (2004) *Assessment using the QCA 'Rules Base' with process-based data*, Advisory Council for New Technologies in Education, Cambridge, 10 February.
- Bennett, Jenkins, Persky & Weiss (2004) *Assessing complex problem-solving performances: the NAEP problem solving in technology-rich environments project*, Advisory Council for New Technologies in Education, Cambridge, 10 February.
- Bennett (2001) How the internet will help large-scale assessment reinvent itself, *Education Policy Analysis Archives*, 9(5).
- Nisbet (2002) *Special arrangements for using ICT in SQA assessments*. Project report to Scottish Qualifications Authority. CALL Centre, University of Edinburgh.
- Nisbet (2003) *Accessibility of SQA assessments in Portable Document Format*. Project report to Scottish Qualifications Authority. CALL Centre, University of Edinburgh
- McAlpine & Ware (2003) Introducing computer assisted assessment in Scotland: laying the foundations for an integrated approach, *29th International Association for Educational Assessment Conference*, Manchester, 5–10 October.
- McCabe (2001) Computer assisted assessment (CAA) of proof = proof of CAA—new approaches to computer assessment for higher level learning, *6th International Conference on Technology in Mathematics Teaching, ICTMT-6*, Klagenfurt, Austria.

McKenna & Bull (2000) Quality assurance of computer assisted assessment: practical and strategic issues, *Quality Assurance in Education*, 8(1), 24–31.

Mitchell, Aldridge, Williamson & Broomhead (2003) Computer based testing of medical knowledge, 7th *International Computer Assisted Assessment Conference*, Loughborough, 8–9 July.