Intelligent Assessment

E-Assessment of Short-Answer Questions

A White Paper

Intelligent Assessment Technologies Ltd.
www.IntelligentAssessment.com

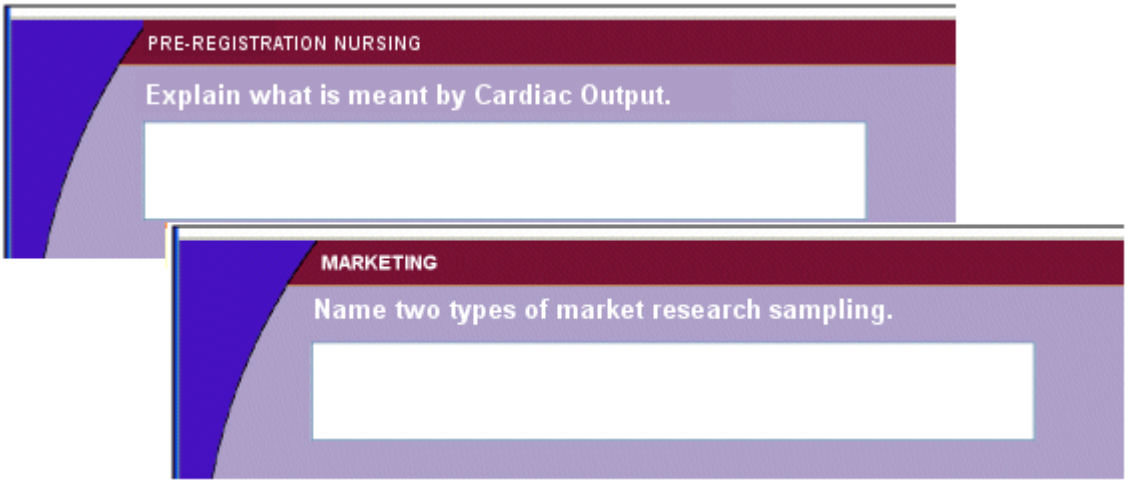
On-Screen Assessment

On-screen assessments are increasingly replacing the traditional pencil and paper tests we all know so well. Not only do the new computerised tests reflect the everyday use of technology, but evidence suggests that they can actually help engage and motivate learners. And from an operational viewpoint of course, on-screen tests can reduce or eliminate the need for human marking, whilst supporting on-demand testing.

The downside of computerised tests is that they rely on closed question types – multiple choice (and variants), drag and drop, image hot-spot etc. Free-text questions are almost completely absent, for the very practical reason that free-text questions are difficult to mark by computer. With the advent of new technology however, some free-text question types can now be automatically marked.

Short-answer free-text questions, a favourite tool of teachers and examiners alike, can now be computer marked by natural language based assessment engines which aim to mimic human marking of free-text. Short-answer questions are traditionally used throughout the learning process because they are believed to reinforce learning and help develop cognitive skills, and they are the preferred instrument of the examiner because they effectively assess understanding without offering prompts or clues.

Figure 1. Short answer free-text questions



But what are the **capabilities and limitations** of computerised marking of short-answer questions ? Before we answer that, we first need to understand a little of how the technology works.

Computerised Marking of Short-Answer Questions

Marking of short-answer questions by computer is analogous to the human marking process. Both the human and the computer will :

- Make use of a pre-defined mark scheme which lists acceptable model answers for each question.
- Analyse each student response in turn to determine if it matches any of the model answers.
- Award marks based on how many, if any, model answers were matched.

Figure 2. Mark schemes for short answer questions are familiar to teachers and examiners everywhere, and list acceptable and unacceptable model answers.

(a) (i) any **one** from 1

- resistant to disease

accept 'having antibodies against the disease'

- will not catch the disease

*do **not** accept 'able to fight the disease'*

(b) can burst the ear drum 1

*accept 'they damage the nerve endings **or** cochlea*
***or** the bones in the ear **or** ear drum'*
*do **not** accept 'they damage the ear'*

But while the overall process performed by human and computer is similar, there are also significant differences.

Computerised marking of short-answer free-text is consistent and efficient, and can provide a detailed audit trail of why marks were awarded. But it is a purely mechanical process, without any subjective capability. And it is not suited to all free-text questions types - the mark scheme being applied must be specific and detailed.

Human markers on the other hand can use professional judgement and inference to mark all manner of free-text questions, from gap-fill to essay. But human markers make human errors, and the process is relatively slow and expensive.

So what we have is a computerised model which has certain advantages over the human model, but which is not nearly so flexible. So under what circumstances can we successfully apply the computerised model ?

When is Computerised Marking of Short-Answer Questions Feasible?

It comes down to question types. The key point to understand is that computerised marking of short-answer questions is a mechanical process which searches for model answers within student responses. So given that we understand this, what kind of questions can and can't be computer marked ? The answer is summarised in the following table.

Figure 3. Characteristics of unsuitable and suitable short-answer questions.

Unsuitable questions

- **X** There are too many ways in which a correct response can be expressed.
- **X** Responses are complex in nature.

Example : Give a brief description of the function of the CPU.

Suitable questions

- ✓ Correct answers can be given in short phrases or simple sentences.
- ✓ Identification of correct and incorrect answers is clear-cut.

Example : In the Java Virtual Machine, when is an object eligible for garbage collection ?

In short, many existing short-answer questions are already suitable for computerised marking. Others, written specifically with the professional human marker in mind, are simply too complex or subjective as they stand. But there are things you can do to make questions more amenable to computerised marking.

Designing Questions for Computerised Marking

The range of questions which can be marked by computer can be increased by careful design of the question. There are two rules of thumb :

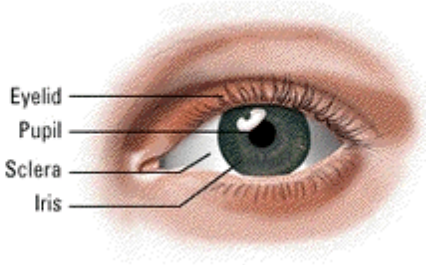
- Word items to constrain students into only writing about one thing at a time.
- Break longer response items down into more specific parts.

Of course this is neither feasible nor desirable in all cases. Nevertheless the opportunity exists to write questions which can be computer marked whilst still maintaining much of the value of short-answer free-text questions.

Figure 4. Designing questions for computerised marking. The first version of the question could not be accurately marked by computer, the second version could.

Version 1

Describe and explain the changes which take place in parts of the eye as light intensity increases.

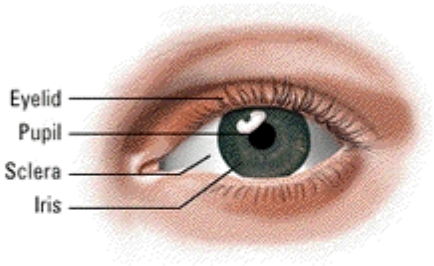


Version 2

What happens to the pupil when the light gets brighter ?

What happens to the iris when the light gets brighter ?

Explain how muscles in the eye cause these changes as the light gets brighter.



The example above shows two versions of a question. They are obviously not exactly equivalent, but one is computer markable to a high accuracy, the other is not. It comes down to whether the compromise is acceptable for your particular assessment.

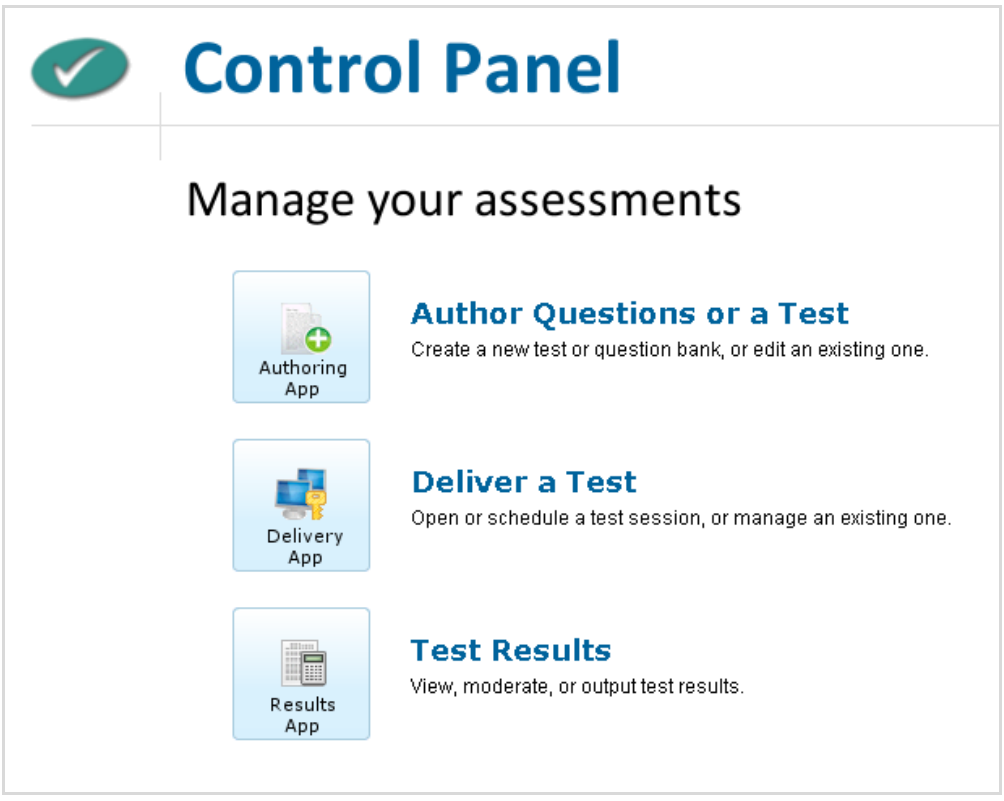
Use of the Technology

So now we know a little about how the technology works, and what it's limitations are, how can it be used ?

Summative Assessment

Integration of a free-text assessment engine with an on-screen examination delivery platform allows the inclusion of automatically marked short-answer questions in summative examinations and tests. Tests are typically marked in batch mode after the test window is closed.

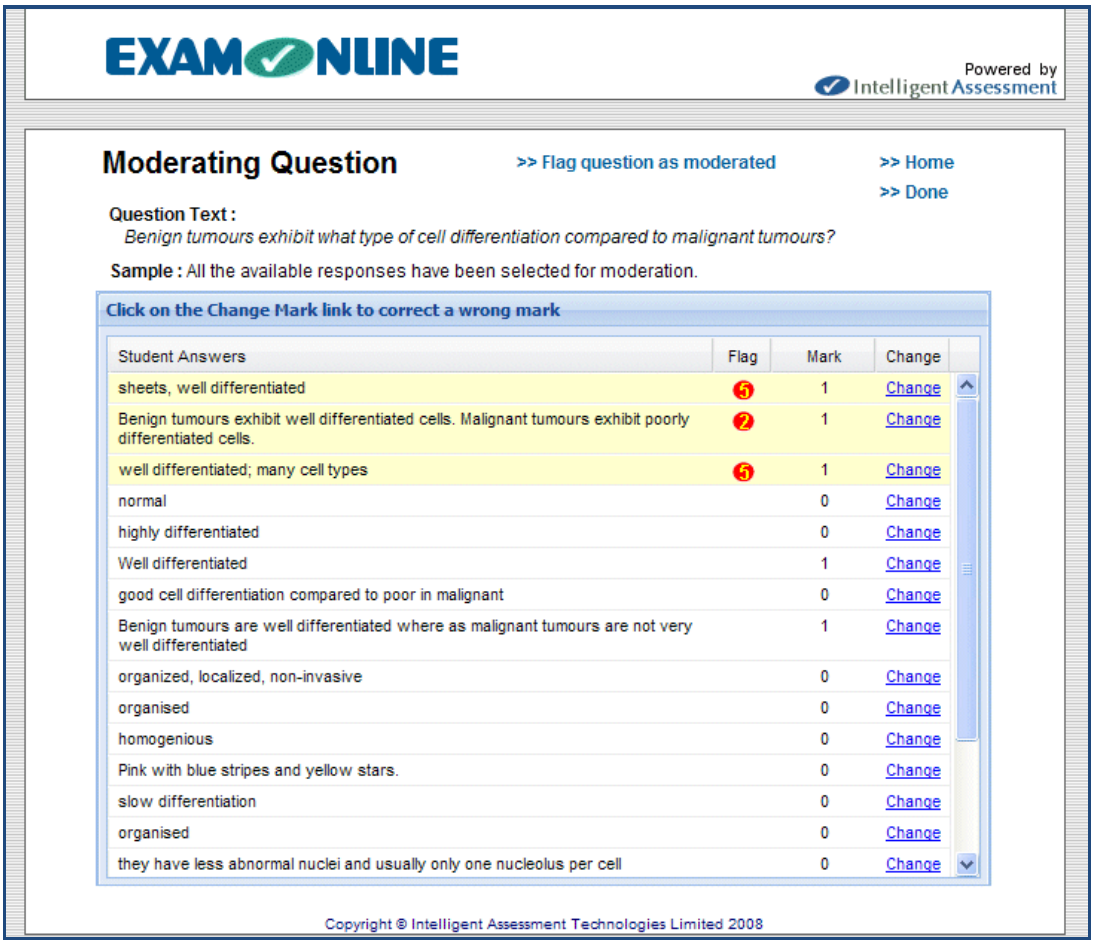
Figure 5. A simple web interface provides access to computerised marking of summative tests. In this example, the tests have been delivered on the ExamOnline eAssessment platform.



For high stakes tests, the marking can be moderated using a sample of the cohort to ensure optimum marking accuracies. This is the model employed by the Medical School at the University of Dundee to test the core knowledge of all medical students. Professor John McEwen, the academic in charge of the computerised test at Dundee, has concluded that the computerised test has a

"very acceptable 'alpha reliability coefficient' of 0.954", and in relation to the other assessment instruments used in the School, "appears to be related best to the written components of finals".

Figure 6. The University of Dundee Medical School employ computerised marking of free-text in on-screen examinations for undergraduate medical students. A web-based interface is used when moderating the questions.



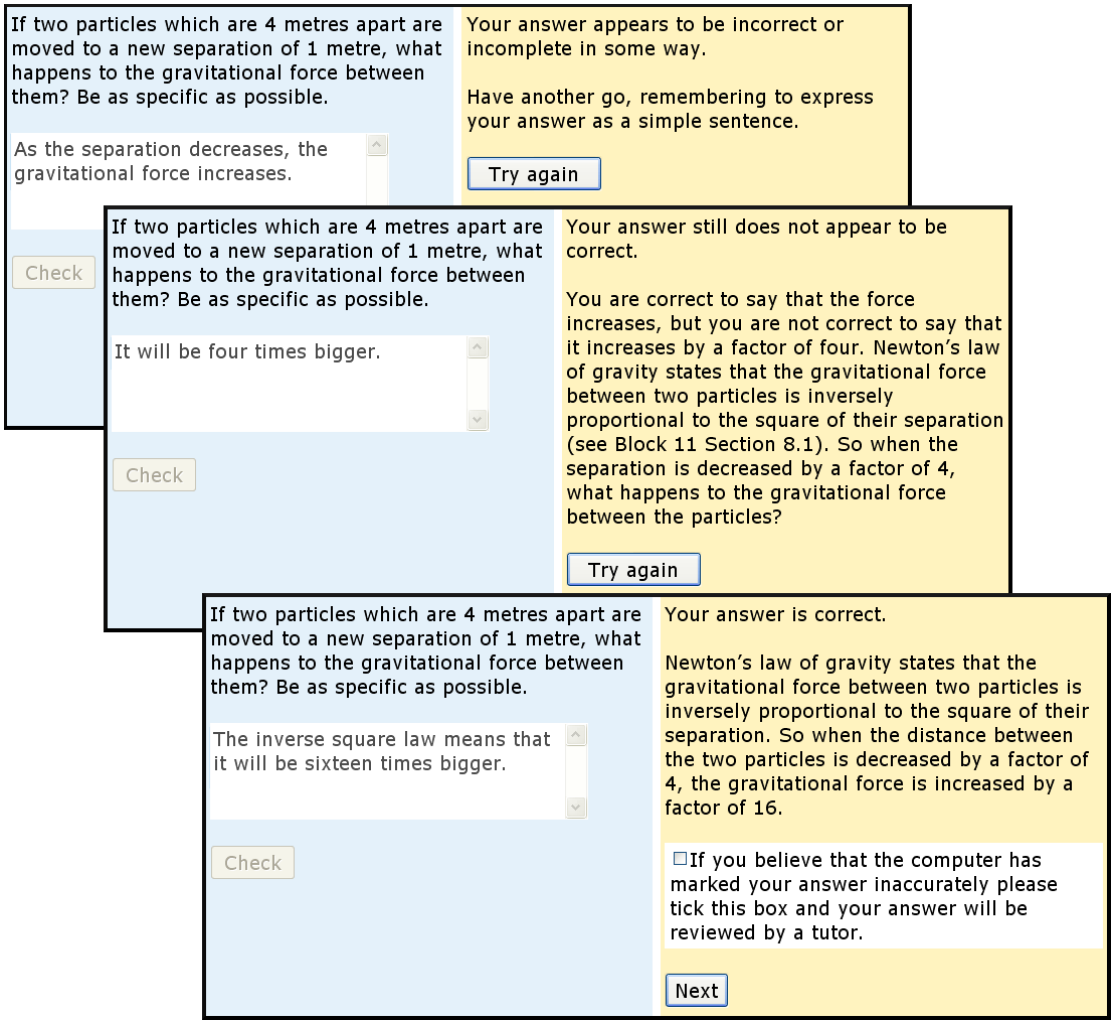
Exam Preparation

Becoming comfortable with the format and question type of the examination is a big part of the exam preparation process. The use of computers to provide exam preparation is now commonplace, but often the format of the preparation tests do not match the format of the exams themselves. With the availability of free-text assessment technology however, the preparation tests can much more closely resemble the real thing for many subjects.

eLearning and Formative Assessment

One of the most powerful uses of the technology may prove to be in formative assessment, reinforcing learning and providing a powerful learning support tool. By configuring the software to look for common misconceptions or partially correct answers as well as correct model answers, learners can receive tailored feedback based on their free-text response.

Figure 7. Providing tailored feedback to pick up on common misconceptions or partially correct responses. This example was created by the Open University.

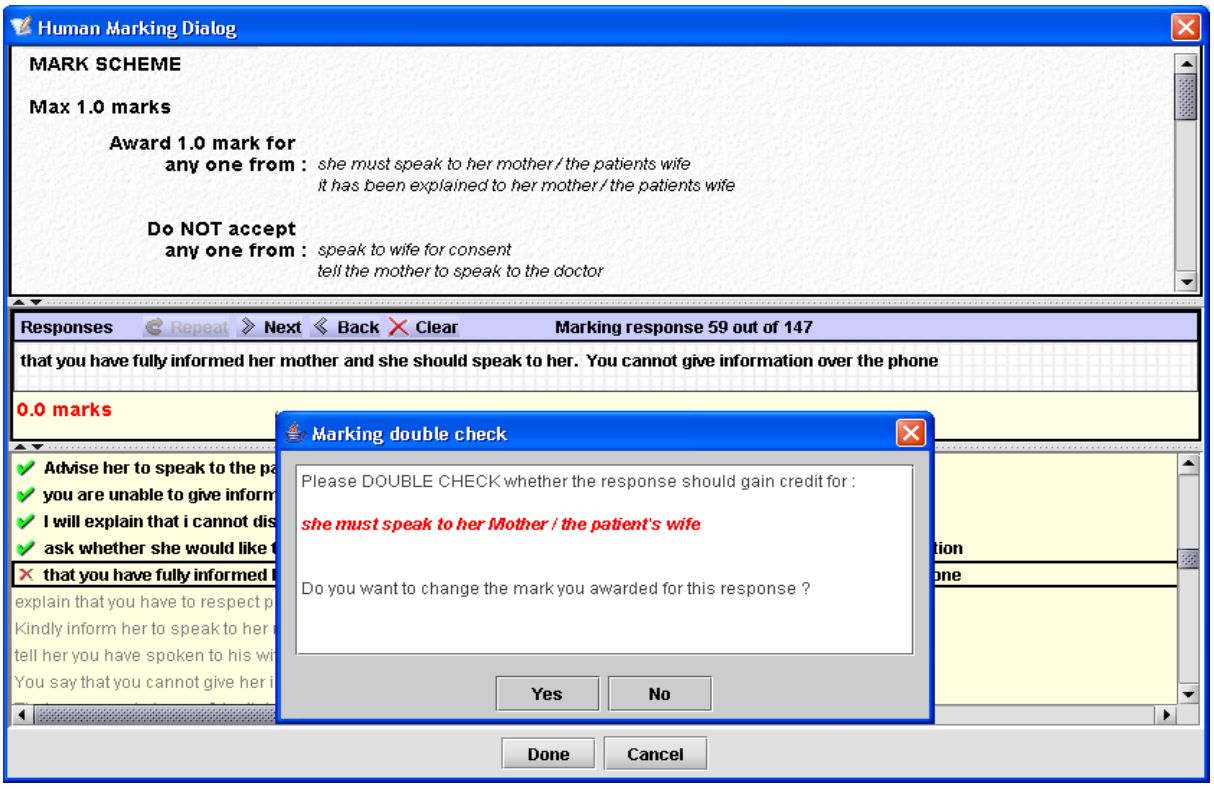


This technique can be very powerful in correcting common misconceptions, in responding to partially correct responses, and in re-enforcing learning. In addition of course, students find this 'conversation' with the computer far more engaging than conventional click-through techniques.

Computer Assisted Marking

Computer assisted marking offers a low risk method of introducing computerised marking of free-text into high stakes on-screen assessment. With computer assisted marking, a human marker can remain as the primary mode of marking, but with each student response they mark also marked by computer. Any discrepancies between the human and the computer mark can be flagged to the human marker for a final decision. In this way the computer can help to detect errors and omissions in human marking which would otherwise go unchecked. The system is also able to collect and collate data relating to the frequency of human marker error.

Figure 8. Computer assisted marking can be used to provide a double check on human markers in high stakes examinations.



Conclusions

Short answer free-text questions are traditionally used in the classroom, in class tests, in revision, and of course in real examinations, but only now can they be embedded in eLearning and eAssessment. This paper has highlighted the strengths and weaknesses of the technology, the ways in which existing free-text questions can be re-modelled to make them computer-markable, and the ways in which the technology can be used.

About Intelligent Assessment Technologies

Intelligent Assessment Technologies (IAT) provide innovative computerised assessment technology and e-assessment consultancy to educational and commercial markets. IAT has pioneered computerised assessment of free-text in the UK. IAT's free-text marking technology has won DTI SMART and SPUR awards for technology.

Contact : **Dr. Tom Mitchell**
01555 660688
tom@intelligentassessment.com
www.intelligentassessment.com
