Intelligent Assessment

E-Assessment of Medical Knowledge

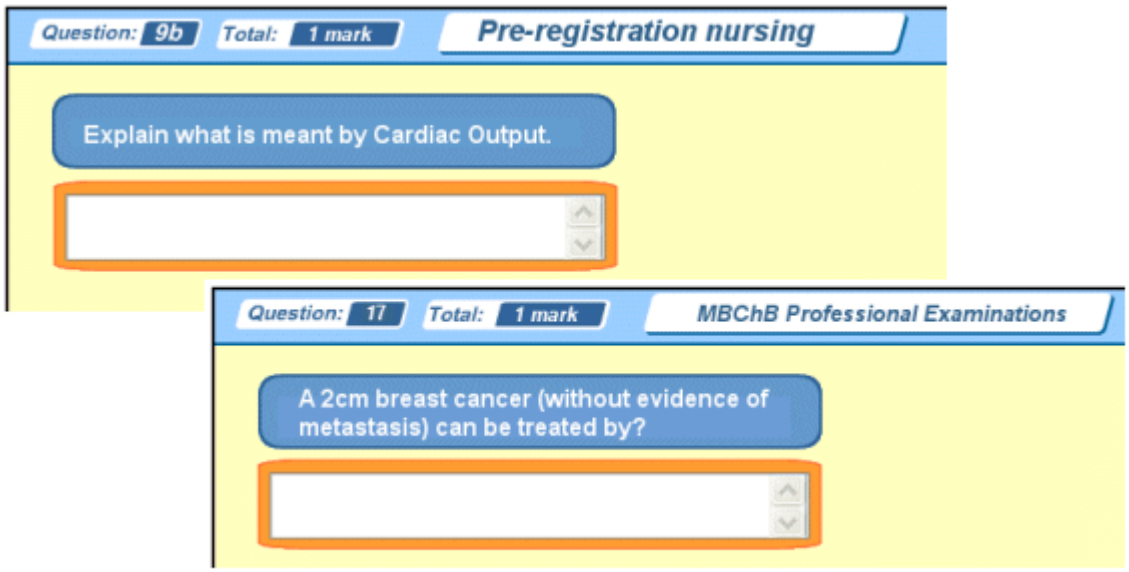
A White Paper

Intelligent Assessment Technologies Ltd.
www.IntelligentAssessment.com

On-Screen Assessment

Assessment of medical, veterinary, or dental knowledge has traditionally relied on **short-answer** free-text questions, because they effectively assess understanding without offering prompts or clues. These questions can now be computer marked by natural language based assessment engines which aim to mimic human marking of free-text.

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.

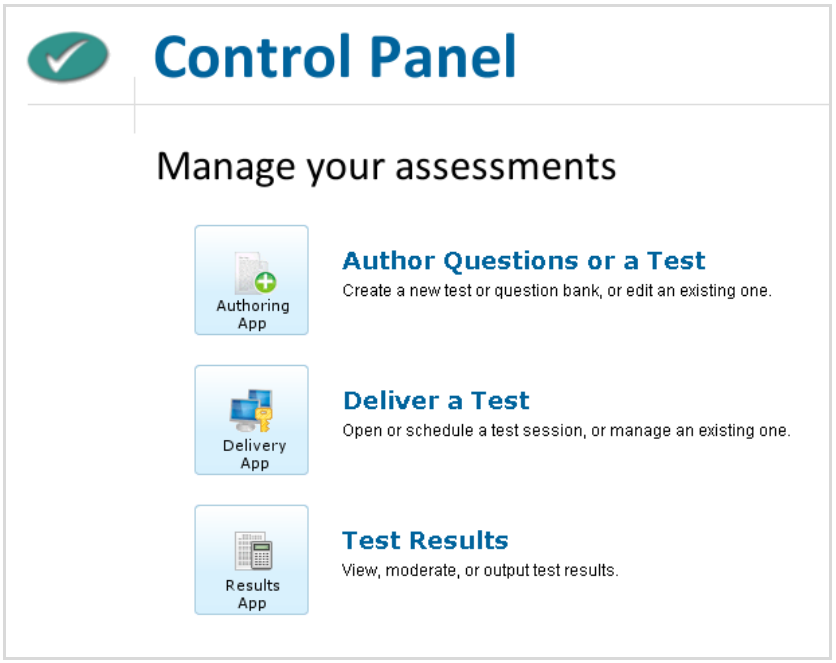
Unsuitable questions	
X There are too many ways in which a correct response can be expressed.	**Example :** Give a brief description of the function of the CPU.
X Responses are complex in nature.	
Suitable questions	
✓ Correct answers can be given in short phrases or simple sentences.	**Example :** In the Java Virtual Machine, when is an object eligible for garbage collection ?
✓ Identification of correct and incorrect answers is clear-cut.	

In short, many existing short-answer questions are already suitable for computerised marking. This is particularly true for assessment of medical knowledge, which typically have a limited range of correct answers, with few 'grey areas'.

Case Study : The Dundee 'Progress Test'

The Medical School at the University of Dundee employ a system supplied by Intelligent Assessment Technologies to test the core knowledge of all medical students. Tests are delivered on-screen using computers in the University IT suite, and students access the test via a standard web browser. The tests are invigilated, and computerised marking of all questions is carried out after test completion. A simple web interface is provided to enable administrators to select which tests to mark, and to initiate the computerised marking process.

Figure 4. Carefully designed interfaces simplify the administration of e-assessment.



As each question in the computerised tests is tagged as to content year, body system, curriculum outcome, core clinical problem and clinical block, the computerised system can provide rich data for feedback to students and teaching staff alike.

Figure 5. Each question in the Dundee Progress Test is tagged, enabling rich formative feedback to be provided to students.

University of Dundee Medical School

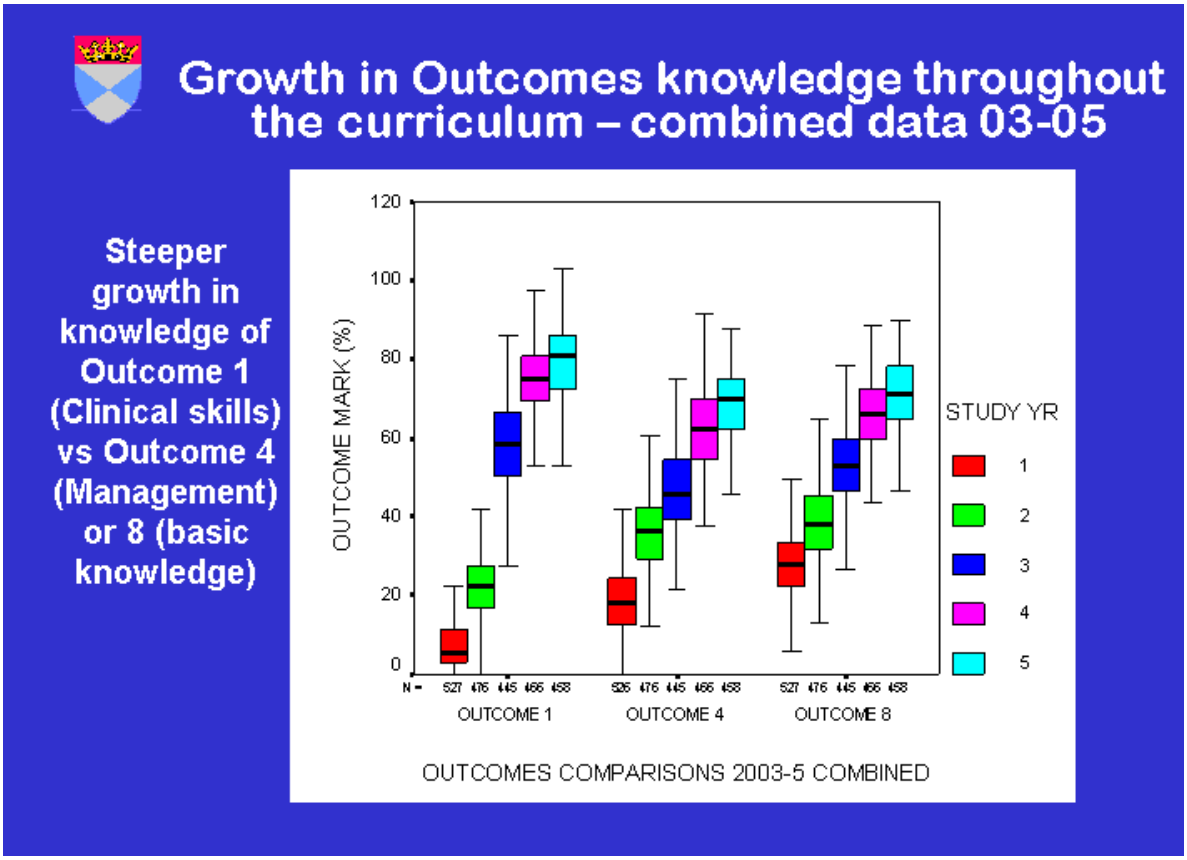
Results of Progress Test

for [redacted]

you where ranked 40th out of 149 3rd year students

Question Group	Max	Your Score	as a %	3rd Year Average
All Questions	270	158	58.5	51.9
By Targetted Year				
1st year questions	56	33	58.9	57.2
2nd year questions	50	28	56.0	43.9
3rd year questions	54	32	59.3	57.3
4th year questions	90	51	56.7	50.9
5th year questions	20	14	70.0	46.9
Outcomes				
Outcome 1	36	23	63.9	51.3
Outcome 2	14	8	57.1	50.2
Outcome 3	29	14	48.3	47.6
Outcome 4	24	15	62.5	52.6
Outcome 5	11	9	81.8	66.2
Outcome 6	2	2	100.0	76.8
Outcome 7	5	3	60.0	45.8
Outcome 8	100	52	52.0	52.7
Outcome 9	12	10	83.3	51.0
Outcome 10	27	16	59.3	51.2
Outcome 11	10	6	60.0	43.8
Systems				
Life Cycle	56	32	57.1	52.3
Renal and Urology	19	7	36.8	53.3
Emergency Medicine	3	3	100.0	65.3
Musculoskeletal	19	14	73.7	52.6
Special Senses	25	13	52.0	55.3
Neurology	24	11	45.8	51.8
Cardiovascular	24	13	54.2	45.7
Gastroenterology	21	12	57.1	55.0
Principles of Health and Diseas	7	3	42.9	46.5
Endocrinology	22	18	81.8	56.4
Haematology	10	5	50.0	35.7
Respiratory	24	15	62.5	53.0
Good Medical Practice	16	12	75.0	49.9
Blocks				
Child Health	9	6	66.7	37.2
General Practice	9	1	11.1	33.6
Medicine	27	16	59.3	56.1
Obstetrics and Gynaecology	9	5	55.6	57.0
Psychiatry	9	5	55.6	52.5
Surgery	27	18	66.7	53.3
Junior House Officer	20	14	70.0	46.9

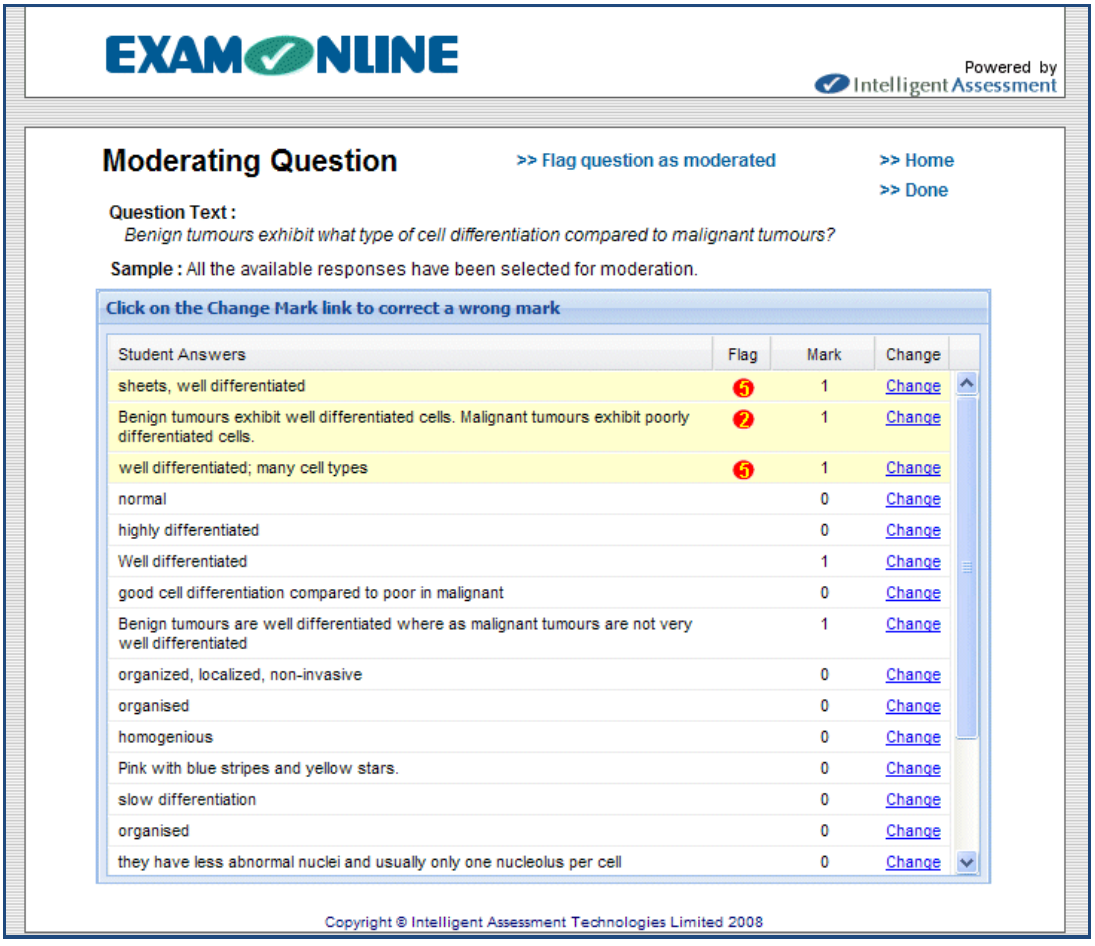
Figure 6. Each question in the Dundee Progress Test is tagged, providing invaluable data to be fed back into the teaching process.



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". The computerised test is now used summatively for Year 5 students, and students transferring in from other institutions must sit the computerised test to assess the level of their medical knowledge against the Dundee learning outcomes.

The Dundee Progress Test won the Summative Assessment Award at the "e-Assessment Scotland: Strategies across Sectors" conference in 2009.

Figure 7. A web-based interface is used when moderating the questions.



Conclusions

Short answer free-text questions are traditionally used in the classroom, in class tests, in revision, and of course in real examinations, and now they can be embedded in medical eLearning and eAssessment.

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