A Comparison of Electronic and Human Marking of Key Feature Examinations in Undergraduate Medical Students

Steve Capey, Adrian Molyneux, Nick Watson & Richard Hays.
Keele University School of Medicine, Newcastle-under-Lyme, Staffordshire, ST5 5BG

s.capey@hfac.keele.ac.uk a.j.molyneux@hfac.keele.ac.uk

Aim

The aim of this project is to develop an electronic, formative, 'key feature' examination and to investigate the feasibility and acceptability of this type of examination with 3rd year medical students.

Background

The development and measurement of clinical decision-making skills has been demonstrated to be essential in undergraduate medicine[1]. The 'key features' approach to assessment has been extensively used in postgraduate medicine to measure clinical decision-making skills[2,3]. The essential element of a key feature question is that only the critical steps in the clinical decision making process are assessed allowing more clinical situations to be tested in a shorter time period than more traditional case based question types. The use of the key features approach can be adapted for use in undergraduate medicine, however they have not been extensively trialled or described in this setting.

Method

A key feature (KF) examination, constructively aligned with the 3rd year of the Validated Keele/Manchester medical curriculum, was developed. The examination consisted of 10 KF stems each with between 4 and 7 questions. The questions and answers were then coded into computer software, (Intelligent Assessment V1.1b). 10 students, recruited into a pilot study, undertook the assessment, under examination conditions at the end of their 3rd year. The students answers were collaboratively marked by two members of the faculty and by the computer software. The computer marking schemes were refined following this and the computer marking was repeated. The individual student scores were collected and subjected to statistical analysis (Students t-test) to identify any significant differences.

Fig 1 A typical 'key feature' question (intermittent claudication)

Problem 5

Lorraine Pepper a 52 year old pottery worker, presented to her general practitioner complaining of pain in her right calf and thigh upon walking 50 yards. The pain started around 6 months ago and is getting progressively worse; it is now impacting her daily activities. She has no significant medical history and smokes on average 25 cigarettes a day. Her GP refers to vascular outpatients. On examination her pulse is 74bpm BP is 140/90mmHg, with an O2 saturation of 99% on air. Physical examination reveals that her feet are cool and pink, she has absent right dorsalis pedis and posterior tibial pulses, her femoral and popliteal pulse are palpable but are decreased in strength. Auscultation reveals a bruit over the femoral artery. Pulses on the left are normal on palpation but a bruit is audible over the femoral artery on that side.

1. What is your most likely diagnosis? Give only one (1)
   i.
2. What are the most appropriate immediate investigations? List no more than (5)
   i.
   ii.
   iii.
   iv.
   v.

Fig 2 Example of the type of image used in questions

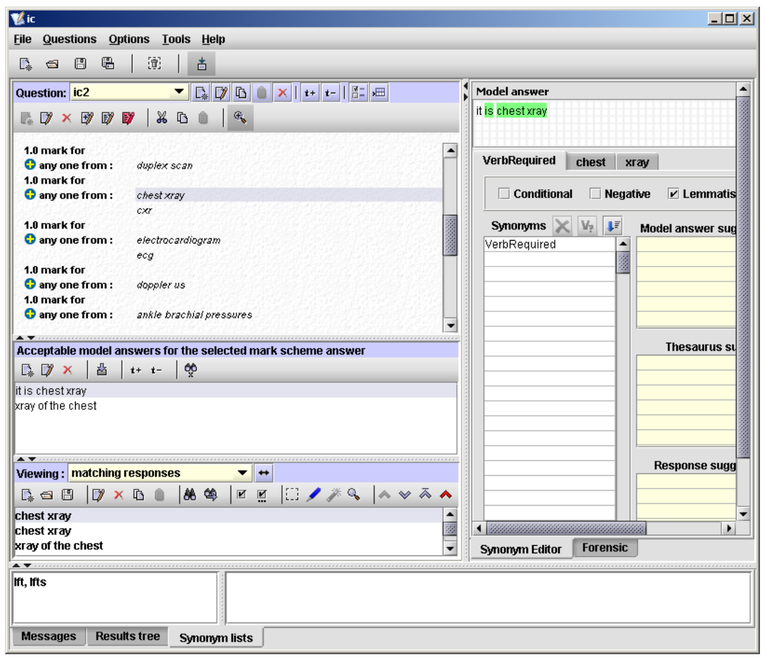

Fig 3 Screenshot of computer interface

Fig 4 Comparison of students scores marked by computer and human

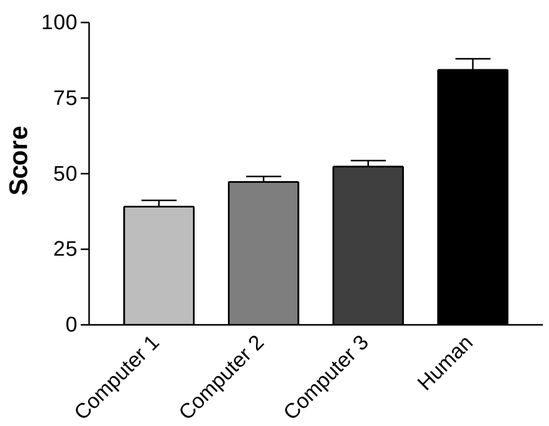

Table 1 Mean total scores

	Raw score	% of human
Comp 1	39	46
Comp 2	47	55
Comp 3	52	62
Human	84	100

Fig 5 Comparison of scores for a typical 'key feature' question for intermittent claudication

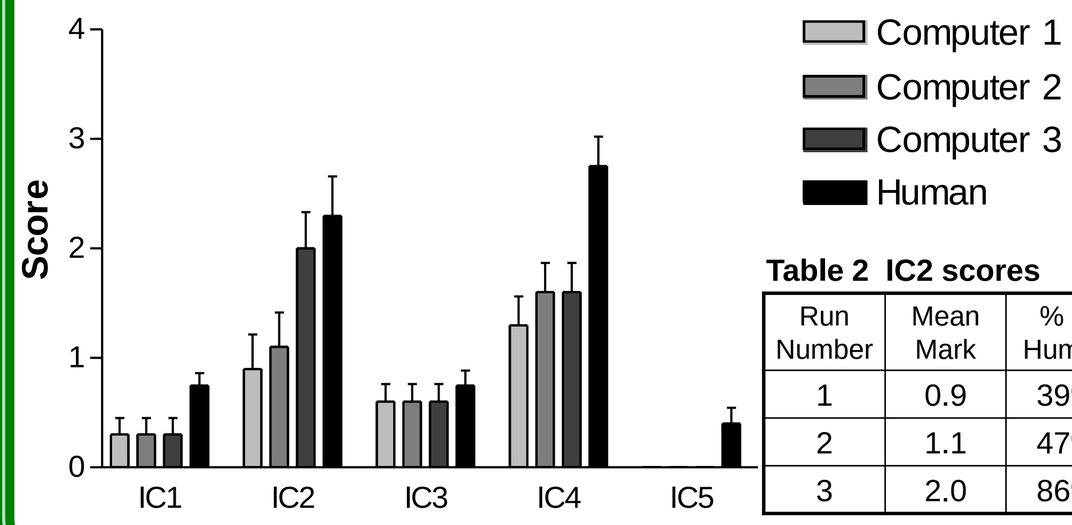

Table 2 IC2 scores

Run Number	Mean Mark	% of Human
1	0.9	39%
2	1.1	47%
3	2.0	86%

Results

The initial marking iteration gave an overall computer - human agreement of 46% across the whole exam. Refining the computer marking scheme in light of student responses to questions increased the agreement to 55%. Finally, the computer mark scheme was further refined to decrease marking errors resulting from unnecessary rigidity in the expectation of inconsequential words, achieving a final agreement of 62%. Even after these 3 runs, the computer mark across the whole test remains significantly different from the human mark. However, some questions (IC2) performed reached a much closer match with human marking and ceased to have a statistical significance (figure 5).

Discussion

This pilot has demonstrated that it is feasible to deliver computer-marked freetext examinations. Previous studies have used a short menu format in computer based key feature examinations[4], however that format is thought to favour weaker students[3]. The difference between computer and human scores remains highly significant precluding its use in summative examination at present. However, it is ideally suited for formative assessment as it provides a robustly tested examination that can be delivered multiple times with minimal marking burden placed on the faculty. It also facilitates the provision of examination practice in a realistic format with the added benefit of providing instant feedback to the students.

References

1 Page G, Bordage G, The Medical Council of Canada's Key Features Project: A more valid Written of Clinical Decision-Making Skills. Acad Med 1995; 104-110
2 Page G, Bordage G, Allen T. Developing Key-feature Problems and Examinations to Assess Clinical Decision-making Skills *Acad Med* 1995; 70(3): 194-201 1995.
3 Farmer E, Page G. A practical guide to assessing clinical decision making skills using the key features approach *Med Ed* 2005; 39:1188-1194
4 Fischer M, Kopp V, Holzer M, Ruderich F, Junger J. A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. Med teach 2005; 5:450-455.