Primum® Computer-based Case Simulations (CCS) for licensing doctors

WORLD CLASS ARENA LIMITED

3 Hampstead West
224 Iverson Road
London
NW6 2HX

+44 (0) 20 7624 1418
+44 (0) 20 7624 1615
info@worldclassarena.net

www.worldclassarena.net

Primum® Computer-based Case Simulations (CCS) for licensing doctors

Contact details

National Board of Medical Examiners (NBME)
3750 Market Street
Philadelphia, PA 19104-3102

Brief details

Primum® is used as one of three parts of the United States Medical Licensing Examination (USMLE®) which is designed to test would-be doctors' ability to treat patients in a practical setting, competencies that were previously examined at the bedside. It is a high-stakes, computer–based case study simulation where candidates are presented with authentic problems and are asked to treat a simulated patient on screen. Candidates, using free text entry, receive information, conduct examinations and order tests and treatments, to which the electronic patient will respond. A candidate's performance is assessed against model responses using a regression-based, automated scoring procedure. The technique is powerful and effective and would appear to be relevant to other subject areas/ professional training courses where "doing skills" are important and systems can be modelled eg economics.

What was the problem?

The United States Medical Licensing Examination™ (USMLE™) is the present three-step examination for medical licensure in the United States and is sponsored by the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners® (NBME). They have been developed from the earlier NBME examinations. In these, until the early 1950s, essays and oral examination were predominant.

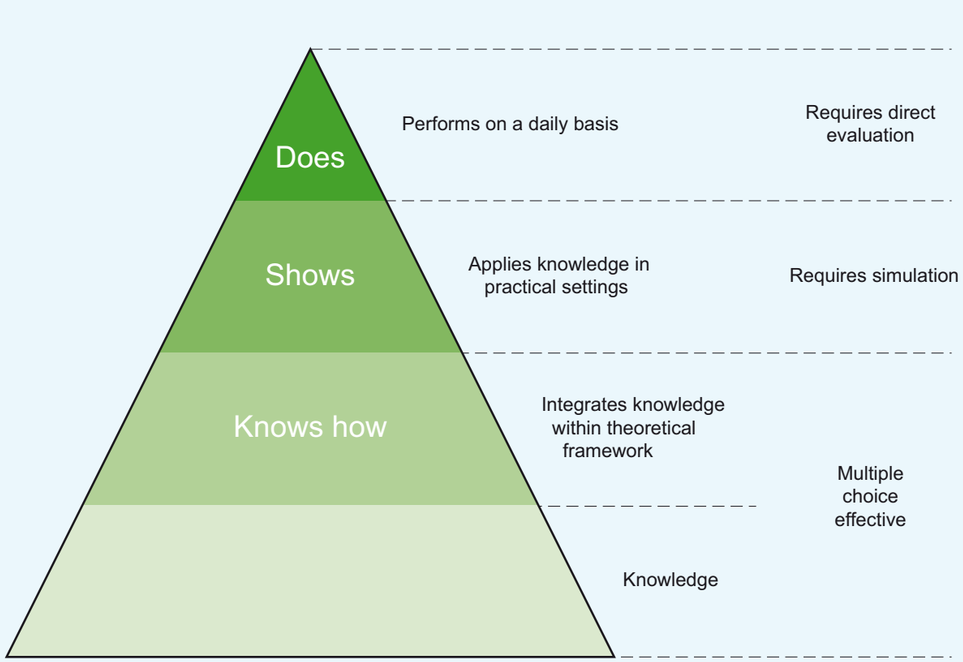
1922- 1950s

Part 1 3 day essay examination in the basic sciences at completion of 2nd year of medical school
Part 2 2 day written examination in the clinical sciences at graduation from medical school
Part 3 1 day practical oral examination on clinical and laboratory problems conducted at the bedside at end of 1st year post graduate.

In the early 1950s, a NBME and Education Testing Services (ETS) study explored the potential advantages of replacing essays with multiple choice tests and concluded that the latter offered greater reliability and validity, conforming more closely with instructor judgements. They were adopted for Parts One and Two. Part Three remained the same.

In the 1960s, concerns about inter-rater reliability and resource implications in terms of both examiners and patients led to the dropping of the Part Three bedside oral component. But it was felt that multiple choice questions reliably addressed lower taxonomic levels but not higher ones and that it was unsatisfactory to rely solely on multiple choice items for the assessment of clinical skills.

For this reason, complex, paper-based Patient Management Problems (PMPs) were developed. Here a candidate's response to a scenario led to a further set of options and thence to a further set and so on. But this was still a matter of selecting rather than constructing an answer and, if the candidate read ahead, s/he would gain illicit insights into the correct answer. Also the scoring was problematic; adequate scores might be obtained by simply avoiding potentially dangerous or overly intrusive actions. Whereas the oral bedside examination had failed on the count of standardisation, the PMPs were too scafffolded, insufficiently open and therefore of limited validity and subject to manipulation.



The solution

In 1999, the whole USMLE® assessment system was computerised, allowing the inclusion of computer-based case studies (CSS) as a new Part Three. Each examinee had to address nine cases (in addition to 500 multiple choice items, tutorials and a questionnaire).

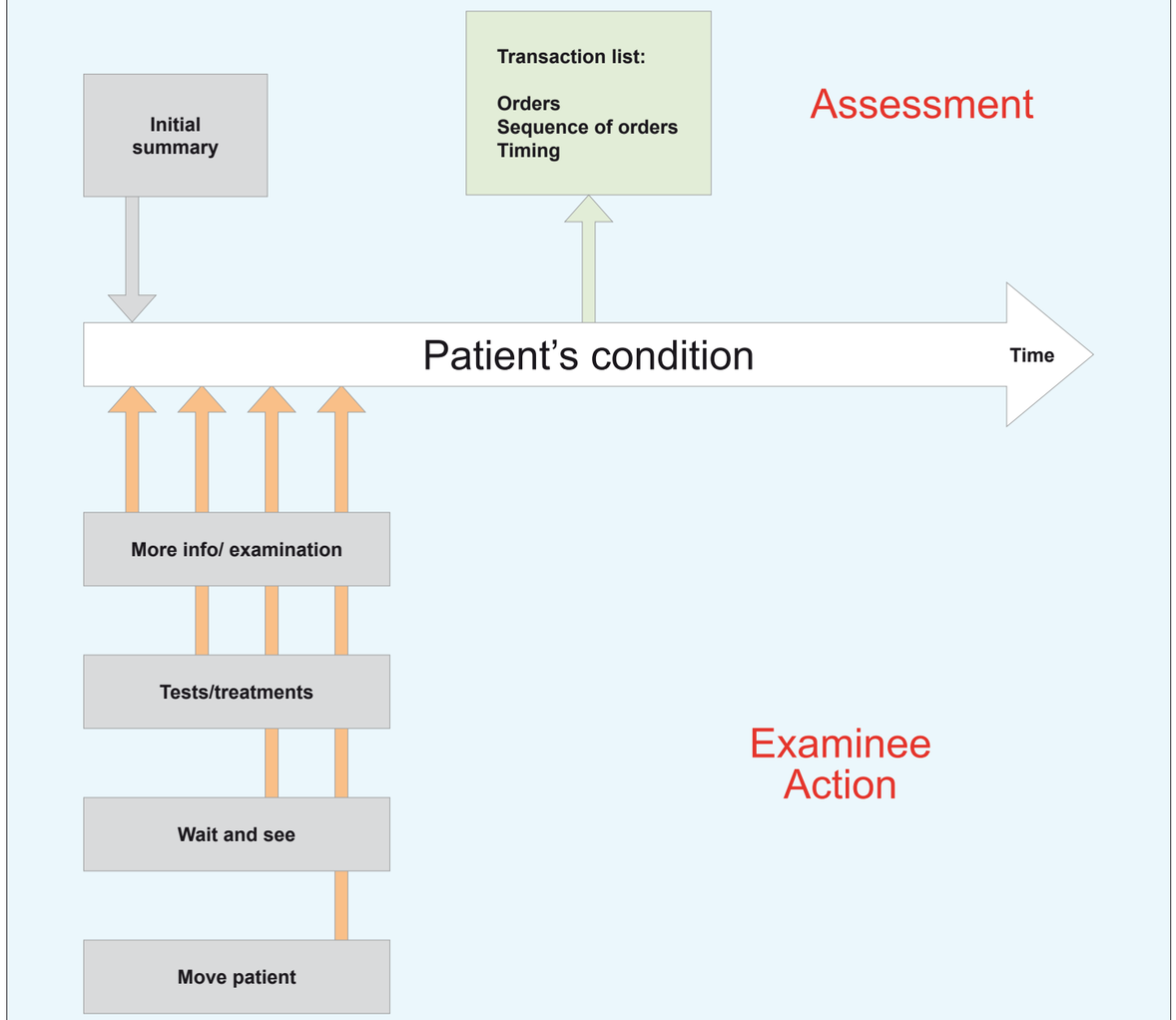
The software presents the candidate with a scenario and a control screen. From here, candidates can

1. request more comprehensive history or physical examinations
2. order tests or procedures through a free text entry order sheet
3. advance the time (to see the results of their actions or inactions)
4. move the patient, perhaps to intensive care or home

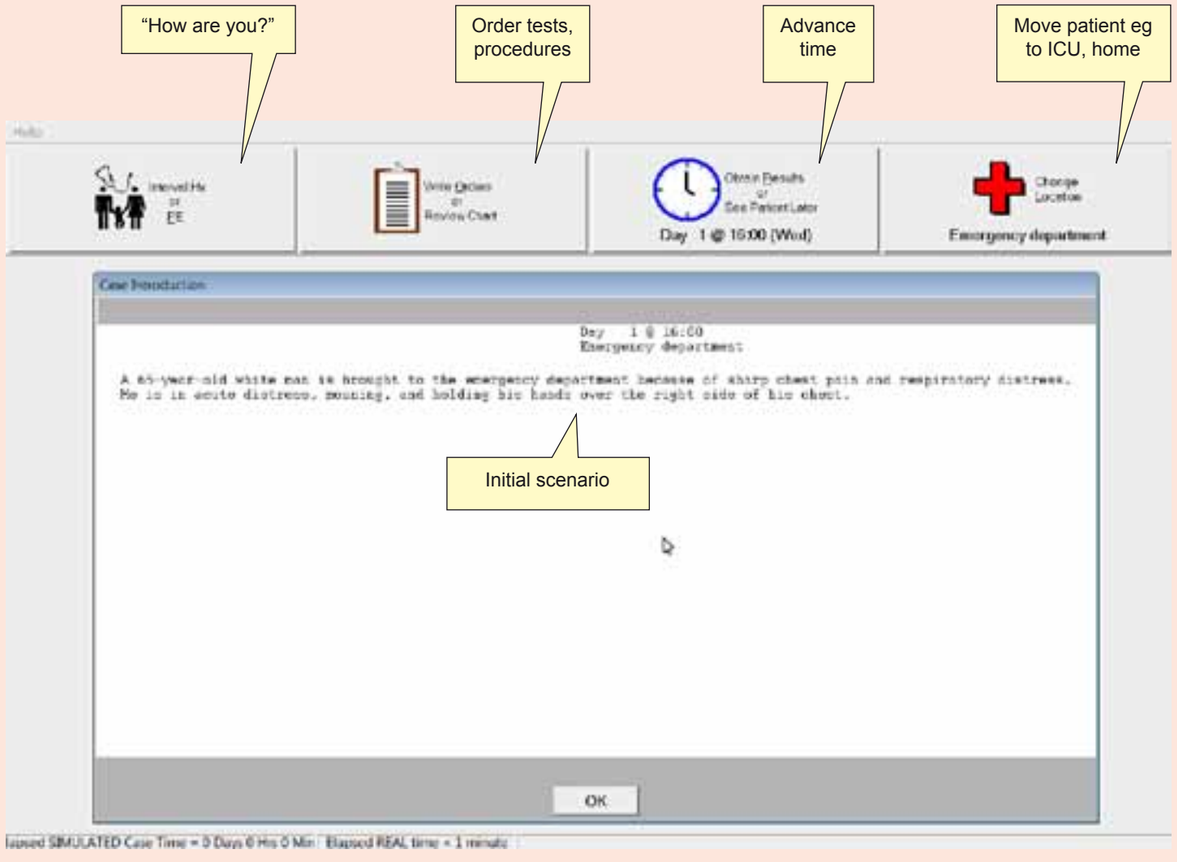
and the candidate is then free to perform whatever actions seem fit. Several thousand tests, treatments and other actions are available. The candidate makes free text entries and the system recognises abbreviations, brand names and acronyms. There is no parser as such and verbs are generally redundant. A three-letter initial sequence in the order is sufficient identify a sequence of options from which the desired action can be selected..

Demonstration software can be downloaded from the USMLE® website http://www.usmle.org/Examinations/practice_materials.html.

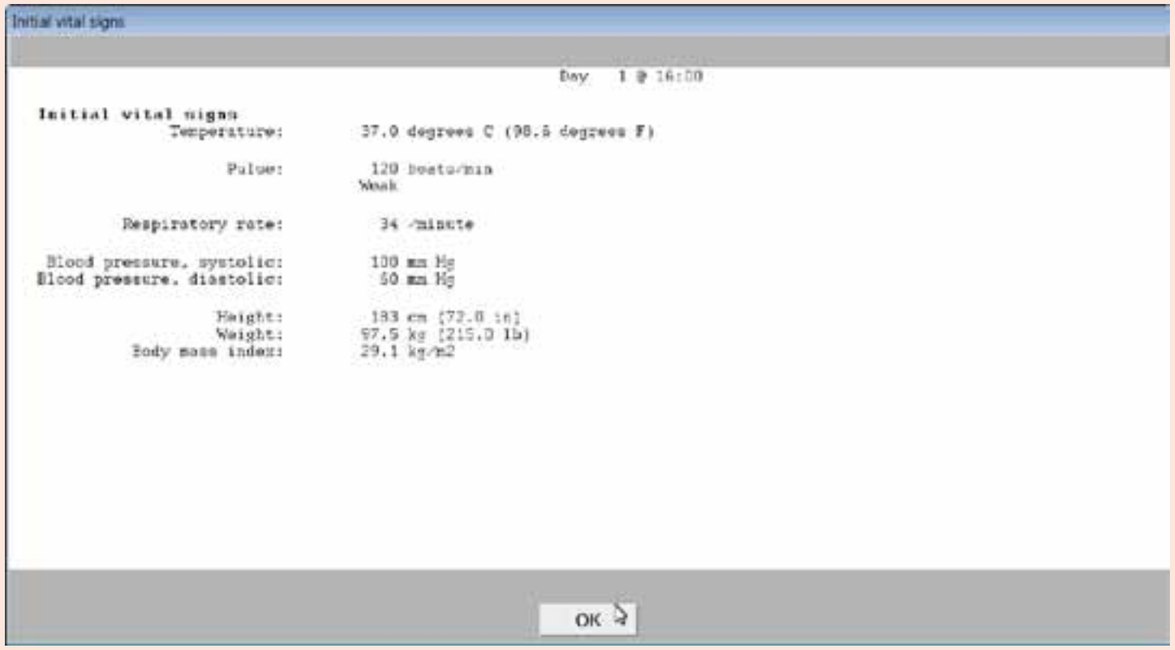
What follows is a brief walk through of case study one.



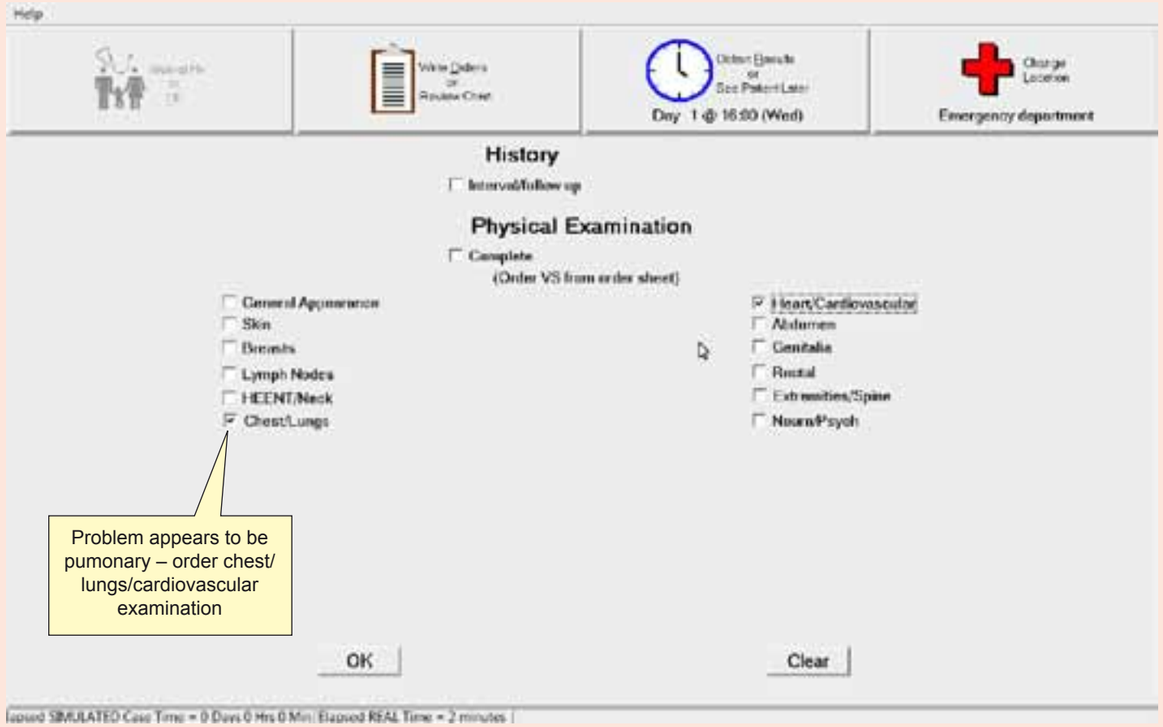
At the start , basic information is provided



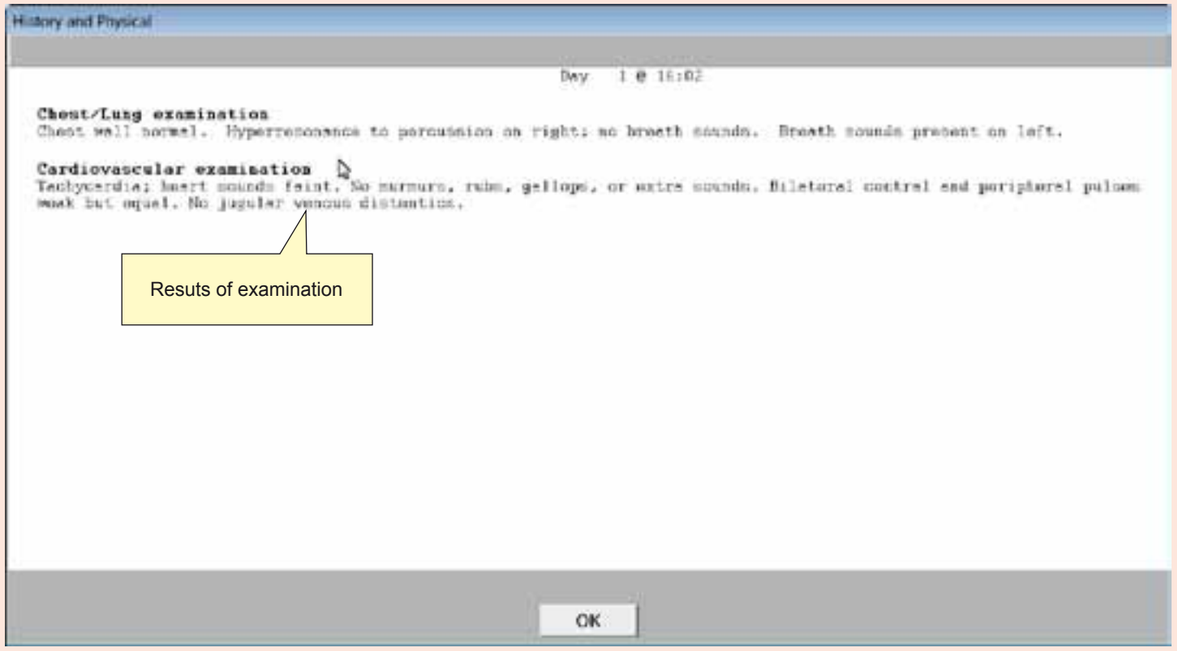
You (the candidate) call up the initial vital signs and a slightly fuller history.



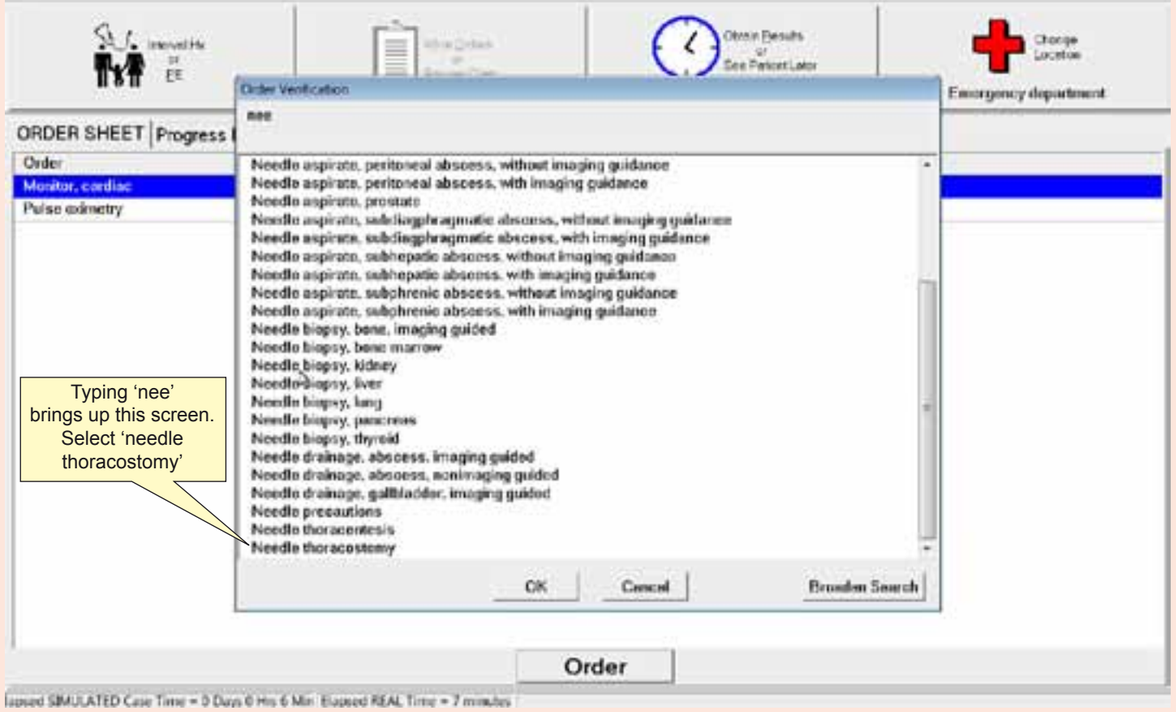
You (the candidate) call up a physical examination:



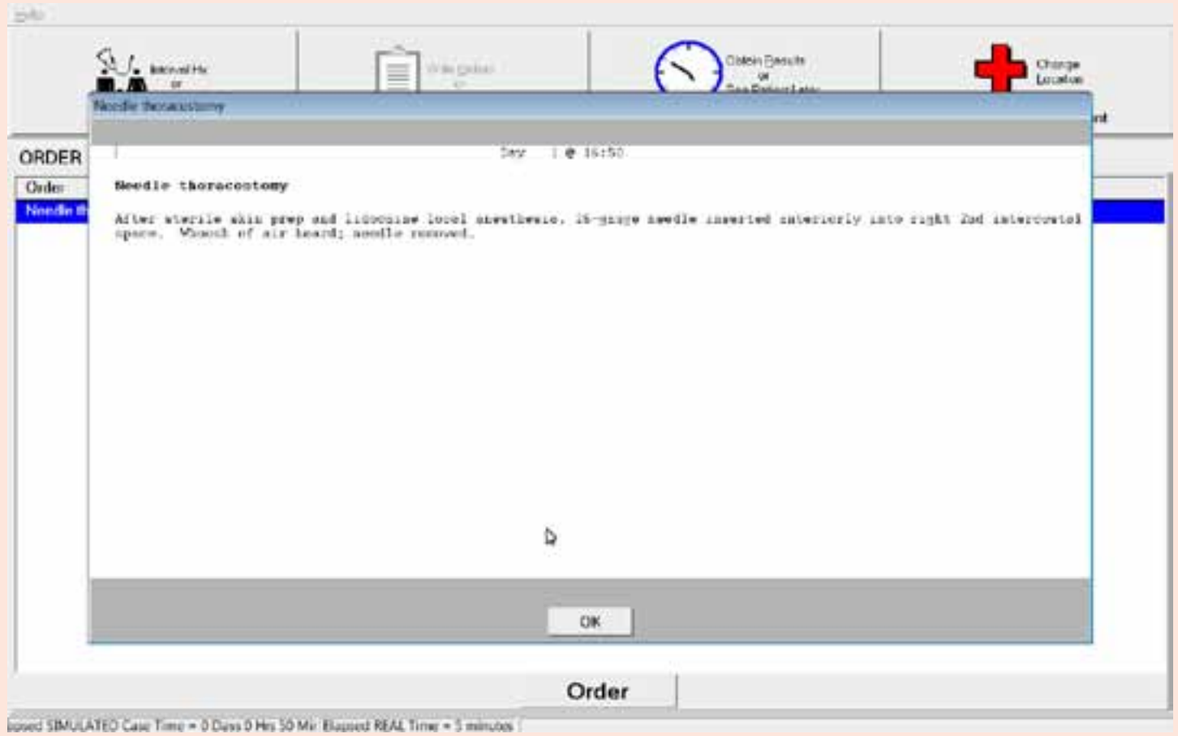
… and check the results.



You decide to order cardiac monitoring, and pulse oximetry to assess oxygen saturation. As soon as the absent breath sounds are discovered, you order a needle thorascostomy followed by a chest tube insertion.



A chest x-ray would need to be ordered to make sure the tube was inserted in the right place and the blood pressure and respiratory rate should be monitored until the patient’s condition has stabilized. The effect of any actions you order will be revealed.



At the end of the case study you are asked for a diagnosis and thanked for looking after the patient (should your treatment be less than optimal, the cases end before the patient actually dies to avoid affecting the examinee in a way that could impact their performance on the next case). The system has stored a transaction list of all your actions, their sequence and timing. It is this transaction list which is scored. (In addition the system stores a complete record of keystrokes that can be used for research purposes.)

Designing a scenario

Primum® uses proprietary software developed by the NBME. The pathways associated with each patient scenario are developed separately and are complex. The number of possibilities can be controlled either by having the case end or by having the patient refuse certain treatments. So for example, the patient might refuse completely inappropriate (and essentially unanticipated) surgery. That requested surgery counts against the examinee in scoring, but doesn't change the patient's condition because it never happened.

Creating each scenario is time consuming and expensive. To be cost effective scenarios must be kept secure and reused.

Scoring and the linear regression model

Experts have to consider the effects of any of thousands of actions and determine whether they are beneficial, neutral or dangerous. Beneficial and dangerous actions need to be rated for degree – eg essential, important or desirable. The appropriateness of actions will also be dependent on timing and sequence. A numerical score is generated based on a linear regression model to produce a score which an expert would have produced. This is done as follows.

Using the software, experts explore the scenario and produce a model answer (including actions and timings) and the associated mark. They then specify beneficial and dangerous actions and associate them with score bands.

These ratings are then tested by the experts, by independently marking sample transaction lists generated by examinees, and discussing their scores. Through an iterative process they achieve common understanding (if not consensus), their scores are averaged and the mean rating used as the dependent measure in a regression equation.

Benefits

Authentic testing was becoming too expensive, too resource-hungry and did not deliver standardised results. The CCS model presents potential doctors with authentic problems where they have to manage the patient in a realistic way. The linear regression model allows judgments to be more consistent than the level achieved through the direct use of experts.
The procedure has been developed over several years and demands continuing appreciable high levels of resourcing in terms of expert panel input. However, it shows considerable resource savings over direct bedside assessment as well as an appreciable increase in standardisation (reliability).

Unlike the PMPs, the computer retains a complete listing of what the candidate has done while managing the simulated case.

It is significant that time is a factor in assessing potential doctors. Where this is not the case, other approaches can be used eg rules based methods (see below).

links and references

Download the *Primum® Computer-Based Case Simulations (CCS) - Tutorial and Practice Cases* at
http://www.usmle.org/Orientation/2009/menu.html

More information is available at:
www. usmle.org
www.usmle.org/Examinations/step3/step3.html
www.usmle.org/Examinations/practice_materials.html

More details of the theory behind Primum® are in A*utomated Scoring of complex tasks in Computer-Based Testing, Williamson Mislevy and Bejar; Mahwah New Jersey 2006,* chapter A regression-based procedure for automated scoring of a complex medical performance assessment, Melissa J. Margolis and Brian E. Clauser.

More details on rule-based methods are in *Williamson Mislevy and Bejar op cit*, chapter Henry Braun et al, Rule-based methods for automatic scoring: application in a licensing context.

Primum® Computer-based Case Simulations (CCS) for licensing doctors