WA4

Certainty-based marking at UCL and Imperial College

A. Gardner-Medwin[1] and N.A. Curtin[2]

[1]Department of Physiology, University College London, London, UK and [2]Division of Biomedical Science, Imperial College London, London, UK

Certainty-based marking (CBM), or Confidence-Based Marking as it was initially termed, was set up in London 12 years ago as a project involving several Physiology Departments. The initial acronym LAPTTOP (London Agreed Protocol for Teaching and Testing of Physiology) soon lost its 'TOP' as usage spread to other areas of the medical curriculum, mainly at UCL and Charing Cross. Current versions of LAPT are on open access (at www.ucl.ac.uk/lapt) and accessed by students from campus computers at more than 30 UK universities. However, despite usually enthusiastic reaction to the concept at conferences, staff at other universities have been slow to engage.

We all want student learning to be more effective and less extravagant in staff time. Part of the strategy can involve self-assessment tasks alongside teaching material, wherever possible challenging deeper knowledge than simply factual or associative learning. A strength of this approach is that staff time can pay off many times over with new student cohorts, but a weakness is that self-assessment is less effective than face-to-face confrontation at probing weaknesses: students who get an answer right often think they knew the answer, when in fact all they did was plump for the more likely answer and strike lucky. A lucky guess is not knowledge, and it is incorrect and inefficient (in statistical terms, adding variance) to mark an assessment as if it were. CBM differentiates between students who may all give the same answers in a test: it rewards those who can distinguish their more reliable and less reliable answers. It places a premium on being able to think through a thorough justification for an answer, and indeed it rewards reflection that leads to a conclusion that the answer is less certain than initially thought. Students find it intuitively easy to use, and cannot cheat by misrepresenting their certainty: correct reporting of one's degree of certainty is always the best strategy to maximise expected score. CBM avoids irrational dilemmas that are often created for students about whether to answer or not, in exercises with fixed negative marking schemes. If you are not familiar with CBM, go to the website (above), try it out, and read the papers linked from the site.

CBM has been used in several contexts at UCL and Imperial. Greatest activity involves online self-assessment exercises designed for teaching and revision (some using past exam papers). Invigilated formative tests with Optical Mark Reader cards (Speedwell Computing Services: www.speedwell.co.uk) have been run with True/False, Single Best Answer (SBA) and Extended Matching (EMQ) question types. Compulsory summative online maths tests (with numerical answers) are employed at UCL, with opportunity for those having difficulty to repeat tests indefinitely with randomised questions and parameters. Online tests are used to establish comprehension of key points following practical work in place of assessed write-ups. In the Imperial clinical course, questions are presented in seminars with students generating self-assessed answers using CBM, with certainty levels linked to qualitative concepts like 'would be willing to proceed on this basis' or 'need to confirm', with enthusiasm from junior doctors about the stimulus it gives to their judgment of clinical knowledge and issues. CBM is used in summative UCL exams with clear evidence of improved statistical reliability and assessment validity with True/False questions (poster communication at this meeting).

Though CBM has been popular and successful, some cautions emerge. Students need practice to use CBM to best advantage, and indeed this is part of the benefit: they learn to judge how reliable is their knowledge. Re-use of published questions in exams becomes a more serious problem: CBM places a premium on good reasons to be sure of an answer, amongst the best of which is of course to have seen the question and answer before. With SBA and EMQ in formative tests, students have tended to be overconfident in their answers. This may result from less practice with these question types, which are less common in our institutions. An alternative hypothesis is that students may narrow options to just 2 or 3 through limited knowledge, and then exaggerate confidence in the final choice because they are correctly confident about some of their rejections. We are keen to discuss evidence bearing on the relative merits and costs of the use of these different question types.

Where applicable, the authors confirm that the experiments described here conform with the Physiological Society ethical requirements.