Optimisation of Certainty-Based Assessment Scores
A.R. Gardner-Medwin, Physiology (NPP), UCL, London

Certainty-Based Marking (CBM)

Aim: To optimise the presentation of
• Self-test Qs so as to challenge students and enhance study
• Exam Qs so as to increase the realistic and predictive value of assessment data

Background:
• Medicine and Physiology often require integration of knowledge from different perspectives to be sure of an answer.
• Thinking about the reliability of knowledge and inference is a key academic skill, with particularly dire consequences in Medicine when it fails.
• Valid measures of knowledge or ignorance must take account of uncertainty.
• Explicit certainty judgment has been shown in many psychological experiments to enhance learning and retention.

CBM is a proper mark scheme in the sense that a student is always motivated and rewarded for distinguishing and identifying those answers that are uncertain and well justified. It is well founded in information theory (see THEORY, to the right).

AN EXAMPLE

Insulin injection raises blood glucose concentration. True/False?

IGNORANCE THEORY

Ignorance has an unambiguous definition:
• Ignorance is a measure of how well the student categorises the certainty of an answer.

CBM gives a good measure of ignorance (= 0.4 x marks lost, relative to correct at C=3):

CBM Implementation: www.ucl.ac.uk/LAPT

All you need to implement CBM for self-tests in your institution (following a model developed for Imperial & Kings, London), are:
• A server site where students & staff authenticate with a local user
• Links to that site, specifying as self-fast, in your VLE
You can use open exercise files, or private ones sited either on your server or at UCL. Editing is simple. Contact me at gardens@ucl.ac.uk for more information. Wholly self-contained software packages are under development, but server loading for new users is almost negligible because computation nearly all takes place on the student’s computer.

SUMMARY

CBM makes sense!
• Doesn’t require special Questions
• Always motivates students to give careful honest judgment

HOW BEST TO PRESENT CBM SCORES

CBM motivates a student to reflect and identify certain vs. reliable answers. This is how you maximise your score.

THE PROBLEM!
This simple comparison, even for students above average at judging uncertainty, can be demoralising and counterproductive. There is nothing wrong with CBM scores; but they are fundamentally different from (and psychometrically superior to) accuracy measures. The problem of presentation is tackled here by generating a “CB Accuracy” by adding a BONUS to the simple accuracy as a measure of how well the student categorises responses as certain or reliable. This bonus is positive or negative, proportional to the amount the average CBM mark is above (or below) the average that would be obtained (shown by heavy black lines below) if the student had used the same optimal C level for all his/her answers. Negative bonuses are common in self-tests when students often have misconceptions (confident errors), but as is evident in exam data, students can aspire to gaining positive bonuses of 2-5%.

How well do students discriminate reliability?

Psychometric optimisation of CB Accuracy

A good measure of the quality of an assessment is how well the score or ranking based on one half of the test (e.g. odd numbered Qs) correlates with that based on the other half.

This correlation is substantially enhanced with CBM. With CB Accuracy the scale of the “Bonus” added to Accuracy is a variable (<0.125 to ensure CBM<100%).

Data from 17 UCL yr 1/2 medical exams are used here to assess the validity and reliability of CB Accuracy and to optimise the bonus factor (0.1 in the graphs presented in the left panel).

Mean correlation coefficient, r, for rankings based on odd & even no. Qs, using CBM ie. actual or bonus modified CBM (where bonus modified CBM = actual CBM + bonus factor x (Accuracy - 1)), as a proportion of the maximum possible correlation is shown.

Enhancement of predictive power of answers (increase of r) compared with right/wrong marking based on the above correlations. A factor of 1.5 means that the benefit is approximately equivalent to a 50% increase in O-ranks in a conventional test.

A bonus factor of 0.1 is clearly a good choice, giving nearly as good psychometric reliability as the CBM mark itself, and better prediction of accuracy and rank based on accuracy on the complementary Q set.