Certainty-Based Marking (CBM) rewards students for identifying which of their answers are uncertain and which reliable. Confident answers incur negative marking if wrong, while uncertain answers receive lower credit but no penalty. The scheme at UCL is shown in Table 1 (1,2). In self-tests this encourages reflection and deeper learning, because the student gains from evaluating both doubts and justifications. CBM gives a sound measure of knowledge (1) and enhances the statistical reliability of assessments (3,4). Problems arise, however, with student perception of CBM final scores, because average marks as a fraction of the maximum are lower than the percentage of correct answers ('accuracy'). For example, students with 80% accuracy typically get only 50% of the CBM maximum (Fig.1).

A new strategy is analysed here using data from 17 medical exams with CBM at UCL (250-300 true/false Qs, 320-360 students). In this, a CB bonus (+ or -) is added to conventional accuracy to reflect how well or badly the student has distinguished uncertain from reliable answers. This bonus is a fraction of the difference between average CB mark and the average that would have resulted with a uniform certainty level, appropriate to the student's overall accuracy. The multiplying 'Bonus Factor' (BF) is shown here to be optimally ca. 0.1 to give maximum statistical reliability for the resultant 'CB Accuracy' (accuracy + bonus). Reliability was calculated as the rank correlation (r) between scores on subsets (odd and even questions) in each exam. This increased from a mean of 0.80 with BF=0 (i.e. conventional accuracy) to 0.86 with BF=0.1 (increase = 0.060 ±0.0039), almost as high as for the CBM average (0.87). This reflects an increase of 57% in predictive power (Fig. 2a), equivalent to adding 57% more questions without CBM. The power to predict conventional accuracy was increased by 16% (Fig. 2b), showing greater validity for CBM scores even with accuracy as the standard.

The CB bonuses added to accuracy in exam data are typically 2-5%, though students doing online self-tests while studying often get negative bonuses due to misconceptions (confident errors). Bonuses and CB accuracy scores are implemented online using BF=0.1 in LAPT at UCL (2), available for use by other institutions. CBM is also available in Moodle, though additional code (5) is required to implement the enhanced scoring and feedback to students.

Acknowledgements: Thanks to D.Bender for providing access to past exam data

(2) *London Agreed Protocol for Teaching (LAPT)*, http://www.ucl.ac.uk/lapt
Table 1

<table>
<thead>
<tr>
<th>Degree of Certainty:</th>
<th>C=1 (low)</th>
<th>C=2 (mid)</th>
<th>C=3 (high)</th>
<th>No Reply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark if Correct:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Penalty if wrong:</td>
<td>0</td>
<td>-2</td>
<td>-6</td>
<td>0</td>
</tr>
<tr>
<td>Probability correct:</td>
<td>&lt;67%</td>
<td>67-80%</td>
<td>&gt;80%</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig. 1 Data from 9,000 online LAPT self-test sessions with >20 answers, showing mean CBM averages plotted against accuracy, for centiles (True/False Qs) or deciles (Single Best Answer, 4-5 options and Extended Matching Questions, 10 options) ranked by accuracy.

Fig. 2 Predictive power \( (r/(1-r)) \) of 'CB Accuracy' scores as a function of bonus factor, calculated from \( r = \text{Pearson rank correlation comparing (a) CB scores vs CB scores and (b) CB scores vs accuracy, for interleaved question subsets. Mean ± sem for 17 exams.} \)