Self-test exercises & exams with Certainty Based Marking (CBM) — Switching to CBM

Tony Gardner-Medwin - Physiology (NPP), UCL

- Judgments in medicine
- The measurement of ignorance
- Worrying features of conventional marking
- CBM: the student perspective
- Enhancement of exam assessment
- Managing self-test tools outside a VLE

Publications, software, try-out, contact, etc: www.ucl.ac.uk/LAPT

The challenge of medicine

The Tricky Bits!!

Medical Knowledge:
- Do I know it for sure?

Medical Teamwork:
- Look it up
- Discuss it
- Refer it

Medical Skills:
- Can I do it well?

Outcome

Judgments in medicine
- The measurement of ignorance

Medical Knowledge
- Turing / Shannon: Ignorance \( \propto -\log(P) \)

With a good mark scheme, ignorance can be inferred in proportion to marks lost

Ignorance (unlike knowledge) has a clear definition

Ignorance from conventional marking (T/F qs)

Firm misconceptions and acknowledged ignorance are treated equally

A lucky guess gets full credit

The most serious inadequacies

(Ignorance inferred in proportion to marks lost)

Ignorance measured with Certainty-Based Marking (T/F qs)

(Ignorance inferred in proportion to marks lost)
Student perspective:
- Always motivated to be honest
- Rewarded for identifying weaknesses
- Rewarded for sound justifications
- Encouraged to reflect & link info
- Misconceptions highlighted
- Simple and transparent scheme
- Perceive it as realistic & fair

Staff perspective:
- Doesn’t require new or different Qs
- Enhanced feedback about content
- Enhanced reliability & validity in exams
- Better student learning experience

An Insulin injection raises blood glucose concentration. True/False?

CBM in Exams
- Standard setters get conventional accuracy (% correct) as well as CBM
- For the same accuracy, % gain if they correctly identify strengths and weaknesses
- CBM is a more soundly based measure of ignorance or knowledge
- CBM yields exam data with greater statistical reliability
- CBM is better than accuracy for predicting accuracy on a separate set of Qs

Data from 1000 random splits of 17 exams (250-300 T/F Qs) into equal subsets:
Correlations are between student rank order on each set, based on Accuracy or CBM

- ↑ of reliability with CBM was equivalent to a 62% ± 7% (sem) ↑ of Q numbers
- ↑ of predictive power for accuracy was equivalent to a 9.2% ± 1.5% (sem) ↑ of Q numbers

CBM Implementation outside UCL (e.g. at Imperial, Kings)

CBM makes sense!
- Doesn’t require special Questions
- Always motivates students to give a careful honest judgement

More sound and fair measure
- ↑ reflection & linking of info
- ↑ realism about uncertainty
- Highlights misconceptions
- Students like it
- ↑ psychometric reliability
- ↑ psychometric validity
- ↓ question numbers
- No loss of conventional exam info

SUMMARY

Tony Gardner-Medwin ucbarg@ucl.ac.uk
www.ucl.ac.uk/lapt
Google: e.g. CBM Medwin UCL Self-tests

How well do students discriminate reliability?
**Additional problems with MCQs (SBA 4 options)**

Only a fraction of extreme ignorance is picked up unless 2nd choices are considered. Total misconceptions and 50:50 with correct answer are treated the same. Full marks for very little knowledge.

(Ignorance inferred in proportion to marks lost)

---

**Ignorance measured with Certainty-Based Marking (SBA 4 options)**

Ignorance:  
- True
- From CBM marks deducted

To achieve these increases using only % correct would have required on average 58% more questions.

---

**Performance in January Formative: first on-paper test in Med Sch**

- Students who did NOT do Self-tests are about twice as likely to fail as students who did Self tests.
- Pattern similar every year: Use is a good predictor of Formative performance.

N.A. Curtis, Imperial College

---

**CBM increases the reliability of exam data with True/False Questions**

'Reliability' indicates to what extent a score measures something about the student's ability, as opposed to 'luck' or chance.

To achieve these increases using only % correct would have required on average 58% more questions.

---

**Knowledge measured with Certainty-Based Marking**

- Luck
- Proper justification requires understanding.

To understand = to link correctly the facts that bear on an issue.

(Technology is how you tell a student from a parrot!)

---

**T/F Qs: Ignorance, as measured with right/wrong Marking (1,0 or 1,-1)**

<table>
<thead>
<tr>
<th>Ignorance</th>
<th>Knowledge</th>
<th>Max. marks</th>
<th>Lucky guess</th>
<th>Avg. mark for guesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Knowledge</td>
<td>1</td>
<td>Ignorance</td>
<td>-1</td>
</tr>
<tr>
<td>False</td>
<td>Knowledge</td>
<td>0</td>
<td>Ignorance</td>
<td>0</td>
</tr>
</tbody>
</table>

Confidence-based marking places greater demands on justification, thereby stimulating understanding.

---

**What is knowledge anyway?**

- Knowledge
- Uncertainty
- Ignorance
- Misconception
- Delusion

Decreasing confidence in what is true, increasing confidence in what is false.

---

**What is understanding?**

To understand = to link correctly the facts that bear on an issue.

(Technology is how you tell a student from a parrot!)

---

**What is knowledge anyway?**

Confidence (Degree of Belief)

Knowledge is justifiable true belief. Proper justification requires understanding.

---

**What is understanding?**

To understand = to link correctly the facts that bear on an issue.

(Technology is how you tell a student from a parrot!)

---

**What is knowledge anyway?**

Confidence (Degree of Belief)

Knowledge is justifiable true belief. Proper justification requires understanding.

---

**What is understanding?**

To understand = to link correctly the facts that bear on an issue.

(Technology is how you tell a student from a parrot!)