UQ – Brisbane – St. Lucia – Jan 22nd, 2014

Certainty Based Marking: Why, How & When?
Tony Gardner-Medwin – UCL – ucgbarg@ucl.ac.uk

• *A lucky guess is not knowledge. A firm misconception is worse than acknowledged ignorance. So why do we mark students as if these things weren't true?* **Ideas, reservations?**

• My motivations? Negative marking? What is knowledge?
• CBM, proper marking schemes, self-tests vs assessments
• CBM: performance in self-tests & exams, CB 'bonus' concept
• Implementation: LAPT, MOODLE, private offline self-test modules

Publications, software, try-out, contact , etc:
www.ucl.ac.uk/lapt   www.TMedwin.net/cbm (new modules)

IDEAS & RESERVATIONS

A lucky guess is not knowledge (T/F?) :
• **T**, but a guess is actually usually informed by <u>some</u> knowledge
• **T**, but on average guesses will give bad marks.

A confident misconception is worse than acknowledged ignorance (T/F?):
• **T** – it can inhibit learning and can be dangerous, but NB misconceptions may reflect genuine knowledge about something related, e.g. "*Australia's capital is Auckland*".

We generally ignore these things (T/F?):
• **T**, but some people think *(incorrectly!)* that negative marking helps by discouraging guessing, or they scale scores so guesses will on average give zero marks.

Why do we ignore them? :
• Conventional marking is simple – any problems will disappear with enough averaging
• Teachers don't really want to discourage answers based on partial knowledge
• Many people dislike negative marking
• People may think that confidence judgements are something separate from knowledge, or not amenable to measurement

---

Irrationality of fixed negative marking

T/F Qs — Fixed negative marking: +/-1

SBA 4 option Qs — Fixed negative marking: +1/-0.33
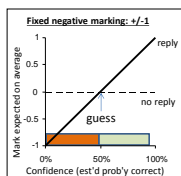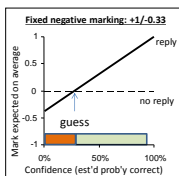
Partial (uncertain) knowledge

It is impossible to hold these values for a preferred option

Students who have the insight to identify which of their answers are unreliable may omit these (perhaps following misguided instruction)

The result is they will on average do worse than students without such insight. This is quite improper and such schemes should be illegal.

MY MOTIVATION

• To help free teachers for what they do best: stimulate interest, creativity, appreciation of deep relationships & corresponding assessments
• Use IT efficiently to supplement teacher activity not replace it
• Use meta-information that is so important in face-to-face assessment
• Encourage student self-tests: practice & challenge, as for sport or music
• Reward the habit of acknowledging uncertainty (in both essays & objective tests)
• Stimulate deeper reflection about Qs (often the same Qs as are already in use)
• Defeat the prejudice that computerised assessment is about rote learning
• Highlight misconceptions - when a student is confident of things that are wrong
• Help students identify weaknesses & strengths, and study accordingly
• Place students more in control of their learning strategy

CBM

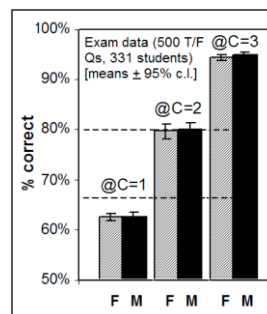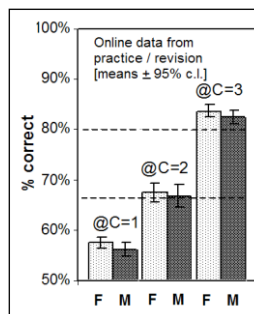| Degree of Certainty :	C=1 (low)	C=2 (mid)	C=3 (high)	No Reply
Mark if correct:	1	2	3	0
Penalty if wrong:	0	- 2	- 6	0
Probability Correct:	<67%	67-80%	>80%	–

Which Certainty Level is Best?
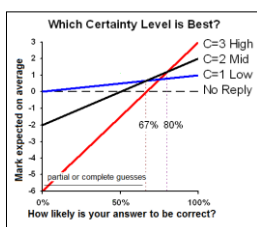
Student perspective:
• Always motivated to be honest
• Rewarded for identifying weaknesses
• Rewarded for sound justifications
• Encouraged to reflect & link info
• Misconceptions highlighted
• Simple and transparent scheme
• Perceive it as realistic & fair
Staff perspective:
• Doesn't require new or different Qs
• Enhanced feedback about content
• Enhanced reliability & validity in exams
• Better student learning experience

How well do students discriminate reliability ?

Online data from practice / revision [means ± 95% c.l.]

Exam data (500 T/F Qs, 331 students) [means ± 95% c.l.]

Slide 1

What is knowledge ?

- ✓ knowledge
- ✓ uncertainty
- 0 ignorance
- × misconception
- ✗ delusion

Decreasing confidence in what is true, Increasing confidence in what is false

Knowledge = justified true belief
Certainty = degree of belief
Justification requires understanding

What is understanding?

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

(How you tell a student from a parrot !)

Nuggets of knowledge

E V I D E N C E → Inference → Certainty (Degree of Belief) → Choice

Networks of understanding

Certainty-Based Marking places greater demands on justification, thereby stimulating understanding

Slide 2

Private Self-Assessment: Why introduce this ?

1. Need for effective tools to supplement staff-student contact time
2. It's good to get students to control and drive their own learning
3. There's increasing scope for self-assessment with IT

The romantic ideal

Assess = ad+sedere =to sit beside

The elements:
Thinking
Challenging
Practising
Correcting
Floundering
Selecting
Discussing
Enjoying

Sports Practice – a model for learning

Challenging
Cooperative
Fun
You learn from mistakes
You mark your boundaries
You push them
Out of view of your coach

Slide 3

CBM Self-tests: what the marks tell you

Very good, but may have repeated self-tests excessively

Good insight into what knowledge is reliable

Underestimates knowledge, or not serious about CBM

Little knowledge but knows what s/he doesn't know

Knows quite a lot but doesn't know where shaky

Misconceptions or lack of awareness of ignorance

CBM mark if you use the same C all the time

NB The CBM mark (as a % of maximum) is always bound to be less than the % correct answers

Slide 4

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

Students who did Self-tests

fail 29%
pass 71%

57% of students did Self-tests

Students who did NOT do Self-tests

fail 56%
pass 44%

43% of students did not do Self-tests

Results for Jan2012

- 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. Curtin, Imperial College

Slide 5

CBM FEEDBACK EXAMPLES (on LAPT: Imperial Self-Test)

Slide 6

Presentation of CBM marks: introduction of a CB 'bonus' concept

Online self-test practice data

repeated (excessive) practice

misconceptions

Exam Data

bonuses mainly +2-5%

Slide 1

CBM in Exams

Best Technology:
Speedwell OMR
Moodle 2.6 (1.9+ with mods)

- Standard setters get conventional accuracy (% correct) as well as CBM
- For the same accuracy, students 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 the 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
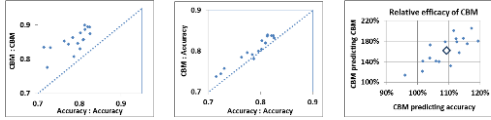
- ↑ of reliability with CBM was **equivalent to a 62% ± 7% (sem)** ↑ **of Q numbers**
- ↑ of predictive power for accuracy was equivalent to a **9.2% ± 1.5% (sem)** ↑ **of Q numbers**

Slide 2

CBM enhances reliability and validity of exam scores
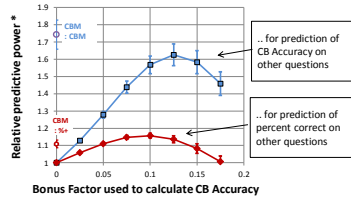
* Factor by which $r/(1-r)$ is increased where r=rank correl. coeff. between scores on odd & even numbered Qs. Mean ± sem for 17 exams, each 250+ t/f Qs, 300+ students.

Slide 3

***Personal self-test software*: download for private practice & learning. Loose linking to an institutional server, LMS**

SERVER

- **Make Comments**
- **Optionally submit results**
- **Access server material**

OFFLINE

COMPUTER / TABLET

- **Download material**
- **Get submitted reports**
- **View Comment Dialogues**

- **Server can provide extra restricted/formative test material & updates**
- **Access via VLE/LMS**
- **Staff editing, data analysis, comments view/respond**
- **Student (wiki) editing & creation of self-tests**

- **Private CBM self-tests**
- **Run offline**
- **Student-centred learning & control**
- **Download from inst'n or public sites**

Info & Download:
www.tmedwin.net/cbm

Slide 4

CBM IMPLEMENTATION

SELF-TESTS
- LAPT (London Agreed Protocol for Teaching)
  Open access & authenticated access from other institutions
- Links from LMS/VLE (BlackBoard at Imperial, Moodle at UCL etc.) to LAPT
- New Self-Test standalone software at TMedwin.net/cbm
  Version for server installation nearing completion
- Moodle CBM core code : now well implemented in Moodle 2.6
  Requires code patches (TMedwin.net/cbm) in Moodle 1.9-2.5

SECURE EXAMS
- Moodle 1.9 -2.6
- Optical Mark Reader Sheets (Speedwell)
- (LAPT & Self-Test modules are designed for fast local feedback and voluntary submission, not secure server-based marking & data collection)

Slide 5

"When you know a thing, to hold that you know it,
when you do not know a thing, to allow that you do not know it
– this is knowledge."   *Confucius*

"... there are known knowns;
... there are known unknowns;
... But there are also unknown unknowns   *Rumsfeld*

"It's not ignorance does so much damage;
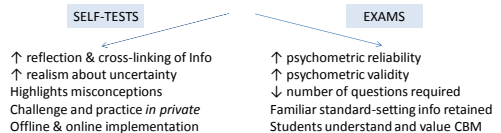- it's knowin' so derned much that ain't so."
attr.: Billings

"A lucky guess is not knowledge.
A firm misconception is worse than acknowledged ignorance.
So why do we mark students as if these things weren't true?"   *TGM*

Slide 6

SUMMARY

www.TMedwin.net/cbm   /   www.ucl.ac.uk/lapt

CBM makes Sense!
Is easily implemented
Doesn't require writing special questions
Always motivates students to identify & acknowledge uncertainty

SELF-TESTS

EXAMS

↑ reflection & cross-linking of Info
↑ realism about uncertainty
Highlights misconceptions
Challenge and practice *in private*
Offline & online implementation

↑ psychometric reliability
↑ psychometric validity
↓ number of questions required
Familiar standard-setting info retained
Students understand and value CBM

Contributors to the project, over many years:
David Bender, Nancy Curtin, Chris Dean, Mike Gahan, Kim Issroff, UCL & Imperial students
Earlier pioneers of work on confidence assessment & learning:
Andrew Ahlgren, Jim Bruno, Robert Ebel, Jack Good, Kate Hevner, Darwin Hunt, Dieudonné Leclercq, Emir Shuford

Contact: Tony Gardner-Medwin – ucgbarg@ucl.ac.uk