

USQ – Toowoomba – Jan 29th, 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? 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

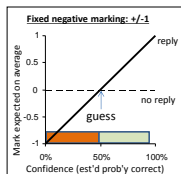
Publications, software, try-out, contact, etc:
www.ucl.ac.uk/lapt www.TMedwin.net/cbm (new modules)

IDEAS & RESERVATIONS

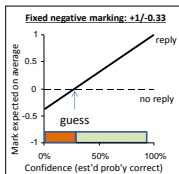
- A lucky guess is not knowledge** NB a guess is usually informed by some knowledge
 NB on average guesses do give bad marks
- A confident misconception is worse than ignorance** It certainly can inhibit learning and can be dangerous.
 But NB misconceptions may reflect genuine knowledge about something related, e.g. "Australia's capital is Auckland".
- We generally ignore these things** NB 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?** Desire to encourage answers based on partial knowledge?
 Dislike of negative marking?
 Averaging is supposed to get round the problems?
 Concern that certainty/uncertainty are different from knowledge?
 or maybe not measurable?

Irrationality of fixed negative marking

T/F Qs



SBA 4 option Qs



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)

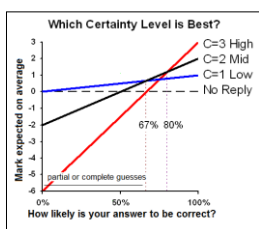
The result is they will on average do worse than students without such insight. This is quite improper and such schemes should be illegal.

MY MOTIVATIONS

- Teachers** → Free them to do what they do best: stimulating *Interest, Creativity, Understanding, Relationships*
 Use IT to help them, not replace them!
- IT** → Use meta-information, as in face-to-face interaction.
 Reward acknowledgement of uncertainty
 Stimulate thinking, not rote learning
- Students** → Encourage control of their own learning
 Learning through mistakes, misconceptions
 Self-challenges in private, like music/sports practice

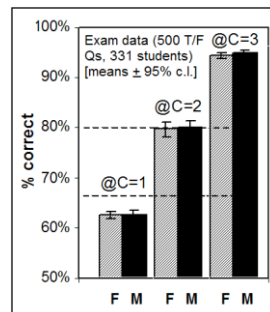
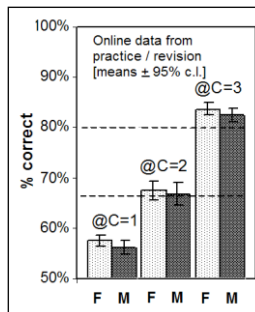
CBM

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%	-



- 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?



What is knowledge ?

- ✓ knowledge
- ✓ uncertainty
- 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 → Networks of understanding → Evidence → Inference → Certainty (Degree of Belief) → Choice

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

Private Self-Tests : Why introduce these ?

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

Music practice

Assess = ad+sedere = to sit beside

The elements:
 Thinking
 Challenging
 Practising
 Correcting
 Floundering
 Selecting
 Discussing
 Enjoying

Sports Practice

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

CBM Self-tests: what the marks tell you

Very good, but may have repeated self-tests excessively

Good insight into which knowledge is reliable

Underestimates knowledge, or not serious about CBM

Knows quite a lot but doesn't know where shaky

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

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

CB Bonus = amount your score is above the black line (for appropriate but uniform C ratings) × 0.1

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

Students who did Self-tests

57% of students did Self-tests

Students who did NOT do Self-tests

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

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

SCORE SUMMARY (exercise=CB3_y1_s2_user=...)

10 responses, 14 correct, CBM Total=111
CB Average= 1.71 mark/Q (max=3), Accuracy= 83%
CB Bonus= 2.3% (giving CB Accuracy = 85%)
 Guesses would give 41% accuracy, so 'Knowledge' = 72% without CBM, and 75% with CBM

CERTAINTY-BASED BREAKDOWN

7 responses at C-1 (4 correct)	57% correct	Target = 67%
25 responses at C-2 (19 correct)	76% correct	Target: 67%-80%
33 responses at C-3 (31 correct)	94% correct	Target: 80%-100%

If your % correct for one of the C levels is shown in colour, then you would have done better setting these C's at a different C level.

YOUR BEST MARK: Your accuracy was in the optimal range for each C level you used. Of courses you can always gain through by better identifying (and up- or down-grading) really uncertain or really reliable answers.

SCORE SUMMARY (exercise=CB3_y1_s2_user=...)

14 responses, 40 correct, CBM Total=111
CB Average= 0.17 mark/Q (max=3), Accuracy= 63%
CB Bonus= -4.5% (giving CB Accuracy = 58%)
 Guesses would give 40% accuracy, so 'Knowledge' = 37% without CBM, and 30% with CBM

CERTAINTY-BASED BREAKDOWN

19 responses at C-1 (10 correct)	53% correct	Target = 67%
4 responses at C-2 (3 correct)	75% correct	Target: 67%-80%
41 responses at C-3 (27 correct)	66% correct	Target: 80%-100%

Since your % correct for one or more C levels is shown in colour out of the optimal range, you would have done better setting these C's at a different C level. This could have gained you 20 extra CB marks, adding 4.7% to your CB bonus and CB Accuracy.

YOUR BEST MARK: Your accuracy was in the optimal range for each C level you used. Of courses you can always gain through by better identifying (and up- or down-grading) really uncertain or really reliable answers.

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%

CBM in Exams

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

	1	A	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

- 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

CBM enhances reliability and validity of exam scores

Relative predictive power*

Bonus Factor used to calculate CB Accuracy

- ... for prediction of CB Accuracy on other questions
- ... for prediction of percent correct on other questions

* 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.

Personal self-test software:

download for private practice & learning. Loose linking to an institutional server, LMS

- Make Comments
- Optionally submit results
- Access server material

- Download material
- Get submitted reports
- View Comment Dialogues

- Private CBM self-testing
- Runs offline
- Student-centred learning & control
- Download from inst'n or public sites

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

Info & Download:
www.tmedwin.net/cbm

CBM IMPLEMENTATION - How?

SELF-TESTS

- LAPT (London Agreed Protocol for Teaching) open access self-tests and you can use your own authenticated access and your own self-tests. Contact me.
- You can link to LAPT from your LMS (BlackBoard at Imperial, Moodle at UCL etc.)
- MOODLE: CBM is now well implemented in Moodle 2.6 (Nov 2013)
Moodle 1.9 – 2.5 require code patches (TMedwin.net/cbm)
- New Self-Test software at TMedwin.net/cbm will replace LAPT, with installation wholly on your own server, plus downloadable personal self-test modules.

SECURE EXAMS*

- Moodle 2.6
- Optical Mark Reader Sheets (Speedwell)

* LAPT & similar software for self-testing is not recommended for secure testing, because feedback is provided locally, rather than from a server.

"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

SUMMARY - Why? When?

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

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

EXAMS

- ↑ 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