Cambridge Assessment
June 2009

UCL

Marks for identifying uncertainty: Stimulation of learning through Certainty-Based Marking

Tony Gardner-Medwin
Physiology, University College London

www.ucl.ac.uk/LAPT

The take home message:

We should reward the acknowledgment of uncertainty

Starting points (you may agree or disagree!)

- The nature of assessment affects how students learn & think
- Objective tests/exercises can stimulate learning & understanding
- Formative assessment is more important than summative

- Different Q types suit different situations, e.g. T/F, SBA, free text
- Scaling to "% above chance" (%Knowledge) should be universal
- *Negative marking* can be either really constructive or really awful

- Students & kids can *enjoy* assessment if it is stimulating, fair, varied, challenging, immediately rewarding, not humiliating -- like a game.

1. ***How Certainty-Based Marking works***
2. ***How it relates to probability & knowledge***
3. ***How students react & use it***
4. ***CBM as summative assessment***
5. ***Why isn't it used more?***

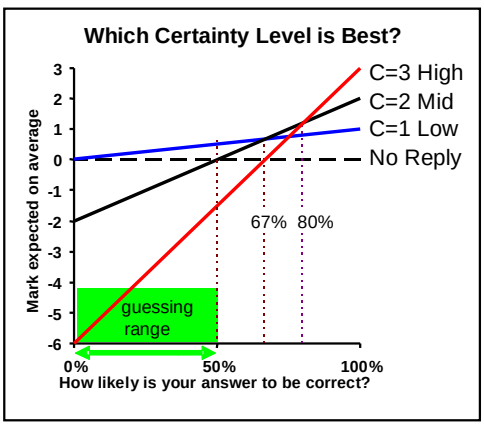
Qu. 3: In a cold winter, the coldest place in a pond is at the bottom.

○ TRUE (or YES)
○ FALSE (or NO)

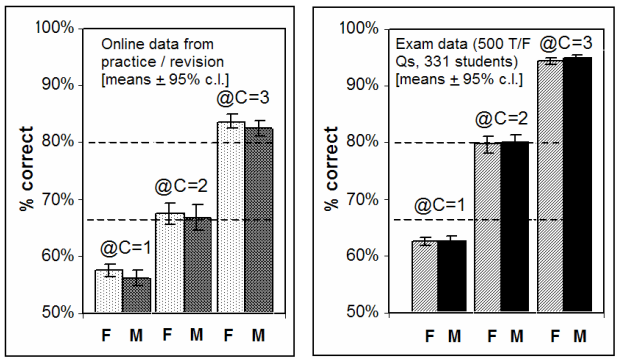
No Reply | **Certainty:** | C=1 (low) | C=2 (mid) | C=3 (high)

Certainty	Low (C=1)	Mid (C=2)	High (C=3)	No reply
Mark if Correct	1	2	3	(0)
Mark if Incorrect	0	-2	-6	(0)

Which Certainty Level is Best?



How well do students discriminate reliability ?



Ordinary words we use to describe Knowledge

✓ knowledge	Decreasing certainty about what is true.
✓ uncertainty	
? don't know	Increasing certainty about something false.
× misconception	
✖ delusion	Increasing "ignorance"

- *Knowledge is a function of certainty (confidence, degree of belief)*
- *There are states a lot worse than acknowledged ignorance*

"It's not ignorance does so much damage - it's knowin' so derned much that ain't so."
attrib J. Billings

"I was gratified to be able to answer promptly, and I did ! - I said I didn't know."
Mark Twain

Student Learning: Principles they readily understand

- You need to know the reliability of your knowledge to use it
- Confident errors are serious, requiring attention to explanations
- Expressing uncertainty when you are uncertain is a **good thing**
- Confidence is about **understanding why things cannot be otherwise**, not about personality
- if over- or under-confident, you must calibrate through practice
- **reflection** and **justification** are essential study habits

In evaluation surveys, a majority of students have always said they like CBM, finding it useful and fair.

They asked to include it in exams, and after 5yrs exam use at UCL they voted 52% : 30% to retain it (in 2005/6), though this was rejected by the conservative medical establishment.

Why test knowledge? Google makes it so easy to find !

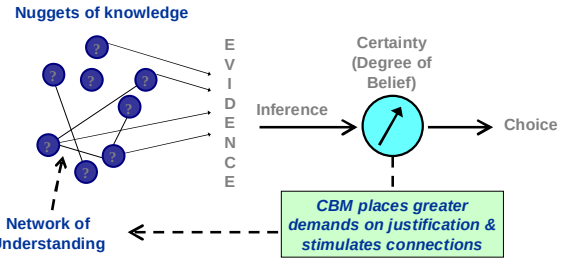
Cheap information (& *increased teamwork*) require :-

1) Identifying things you will get wrong and not Google!
"**unknown unknowns**" rather than **"don't knows"**
2) Judging reliability and uncertainty correctly
.... setting a threshold for seeking help
.... evaluating conflicting and corroborating information

In olden times, you had to rely on your own stored information
.... you would make a best choice and "**go for it**"
School leavers have more sparse (though broader) stored info, but still have a "go for it" culture - to a scary extent!
.... *responding with an immediate idea & not thinking much*

These lessons are core things that CBM teaches

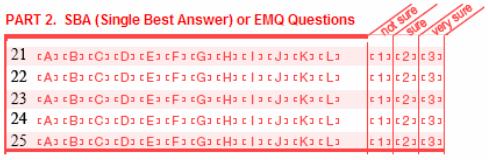
Thinking about uncertainty / justification develops understanding of relationships



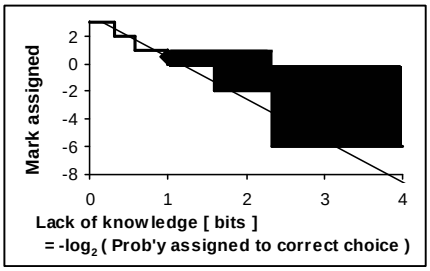
To understand = to link correctly the facts that bear on an issue.

Using CBM

1. With UCL LAPT software, online or from CD
2. With Moodle - work in progress
3. With commercial software – some progress, more needed!
4. Secure exams, with OMR Cards [Speedwell]

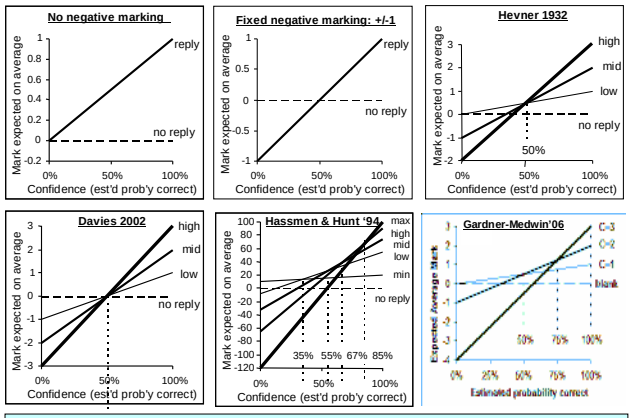


CBM quite closely follows the ideal ignorance measure



The student loses about 3 marks per 'bit' of ignorance - up to a maximum of 3 bits

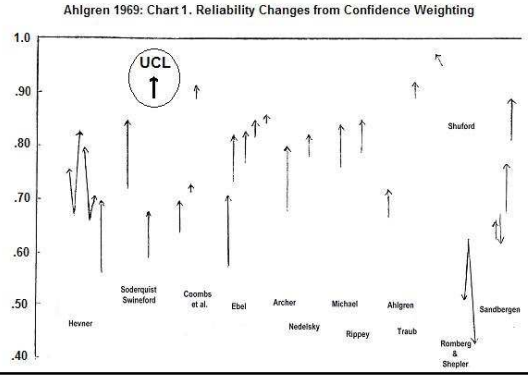
What's a good mark scheme?



The standard LAPT (1,2,3 / 0,-2,-6) scheme seems better than any of these.

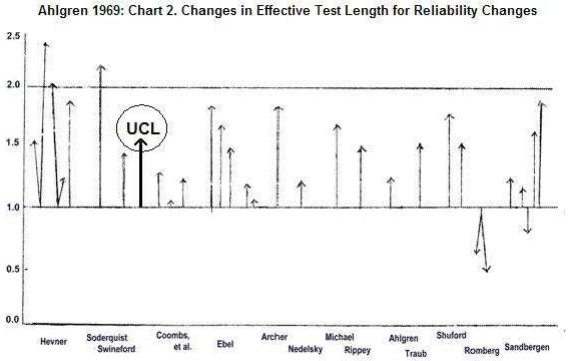
CBM increases the reliability of exam data

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



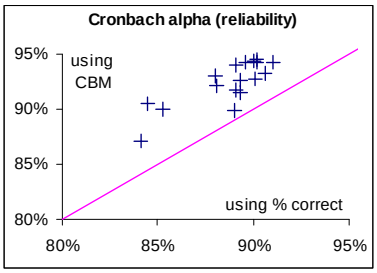
CBM increases the effective test length

With increased 'Reliability' you don't need so many exam questions to get data of equal quality.



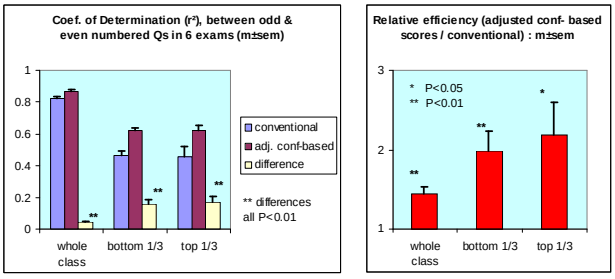
CBM increases the reliability of exam data with True/False Questions

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



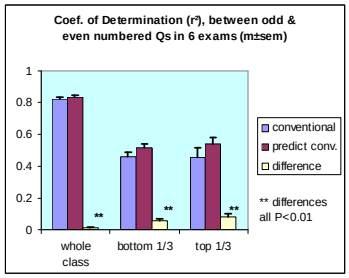
To achieve these increases using only % correct would have required on average 58% more questions.

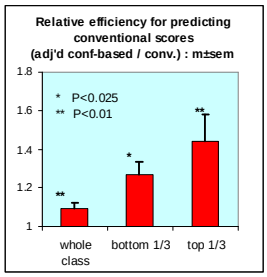
Reliability and efficiency of exams (Quality of data / number of questions) are enhanced with CBM



Data from 6 medical student exams (250-300 T/F Qs each, >300 students).

Certainty-based scores predict the conventional score on different Qs better than conventional scores do.





How should one handle students with poor calibration?

Significantly overconfident in exam: **2 students (1%)**

e.g. 50% correct @C=1, 59%@C=2, 73%@C=3

Significantly underconfident in exam: **41 students (14%)**

e.g. 83% correct @C=1, 89%@C=2, 99%@C=3

Maybe one shouldn't penalise such students

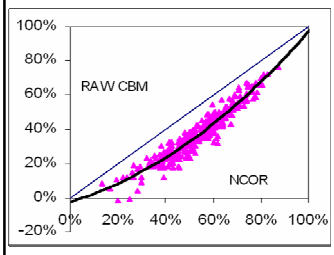
Adjusted confidence-based score:

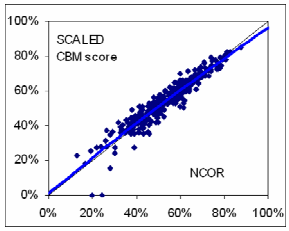
Mark the set of answers at each C level as if they were entered at the C level that gives the highest score**.

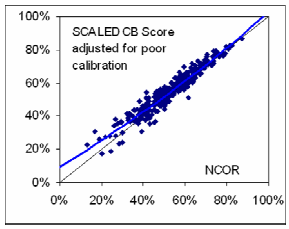
mean benefit = 1.5% ± 2.1% (median 0.6%)

** (first combining sets if %correct is not in ascending order)

Scaling CBM scores to be directly comparable with conventional scores

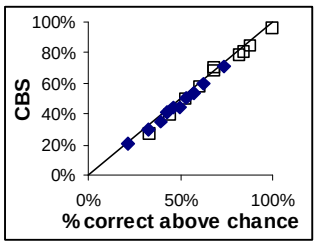






NCOR is based on number correct, scaled so guesses (50% prob'y correct) give on average 0%. ("% Knowledge")

Equivalence of **scaled CBM scores and conventional scores for standard setting.



True/False ♦ and SBA □ (5 option) components of a formative test for 345 students were ranked by conventional scores. Then for each decile, mean CBS scores are plotted against % correct above chance ("% knowledge").

Gardner-Medwin & Curtin 2007 REAP conference, data from Imperial College
****CBS = ((Total-Chance)/(Max-Chance))^p × 100%, where p = 0.6 for TF, 0.48 for SBA (5opt)**

Why doesn't everybody already use CBM ? - a puzzle

- Enthusiasm was exhausted before the age of 'online'
- Some CBM methods were complex, opaque or non-motivating
- Reluctance to treat certainty as integral to knowledge
- Mistaken worries about 'personality bias'
- Under-rating of self-assessment & practice as learning tools
- Worry that CBM would need new questions
- Worry that CBM would upset standard-setting
- Inertia and vested interests

A few of the names associated with confidence testing in education

- Andrew Ahlgren
- Jim Bruno
- Confucius
- Robert Ebel
- Jack Good
- Kate Hevner
- **Darwin Hunt**
- **Dieudonné Leclercq**
- Emir Shuford

London Colleagues:

- Mike Gahan
- David Bender
- Nancy Curtin

"When you know a thing,
to hold that you know it.
And when you do not know a thing,
to allow that you do not know it.
This is knowledge."

"Learning without thought is a waste of time."
Confucius

We fail if we mark a lucky guess as if it were knowledge.

We fail if we mark misconceptions as no worse than ignorance.

www.ucl.ac.uk/lapt