Confidence-Based Marking: The proper strategy for exams

Tony Gardner-Medwin, Physiology, UCL
www.ucl.ac.uk/lapt

- What is CBM ?
- What's it like to experience it?
- Students gain by justifying either reservation or confidence – by thinking more
- CBM is no more subjective than right/wrong
- Exam data is more reliable and more valid
- CBM helps CAL/CAA relieve teachers for roles where they are essential

Assessment for Learning

Confidence-Based Marking scheme

… for objectively right/wrong answers, e.g. T/F, MCQ, EMQs, Numerical

Confidence Level	**1**	**2**	**3**
Score if Correct	**1**	**2**	**3**
Score if incorrect	**0**	**-2**	**-6**
When to use:			
Probability correct	**< 67%**	**>67%**	**>80%**
Odds	**< 2:1**	**>2:1**	**>4:1**

What reasons can you see to prefer or reject such a scheme?

(a) In formative work

(b) in exams ?

The council puts a free rubbish skip in our road on the Sunday after the 2nd Saturday in each month.

This is usually the 3rd Sunday in the month. T/F?

Here's the problem! Many students don't take an argument this far:

- they are not used to trying to justify their answers
- little practice at establishing when an argument is rigorous
- usually can pass exams without this –hunches are good enough !

NB There are a lot of questions like this in the BMAT test available for practice under LAPT at: www.ucl.ac.uk/lapt

Please don't call out the answer !!

Su M …. Sa	Su M …. Sa	Su M …. Sa	Su M …. Sa

Put your hand up!

2nd Sa

1st day of month

falls in this week

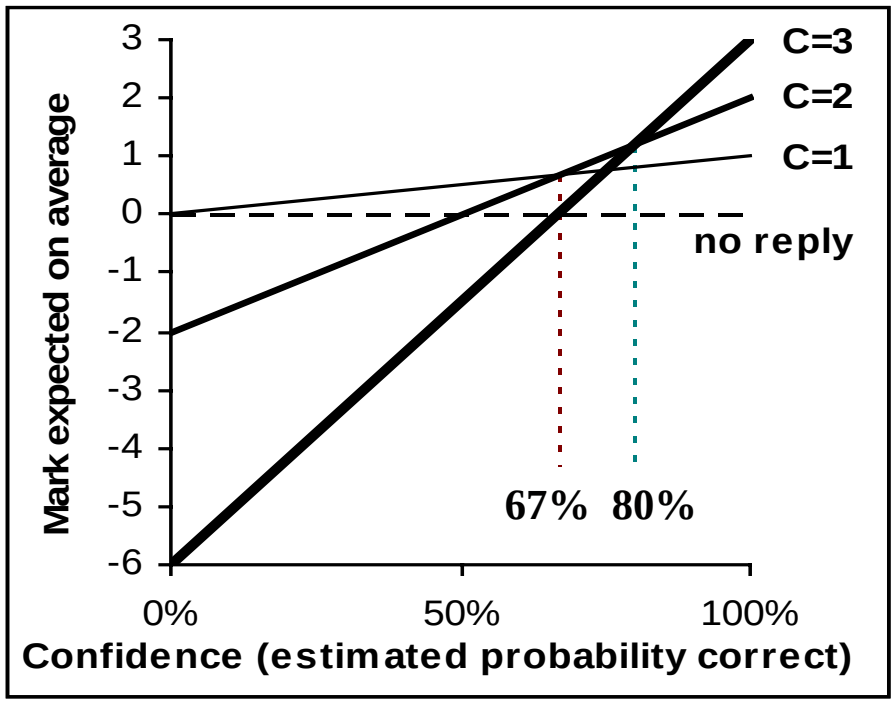
Skip!

With CBM you must think about justification

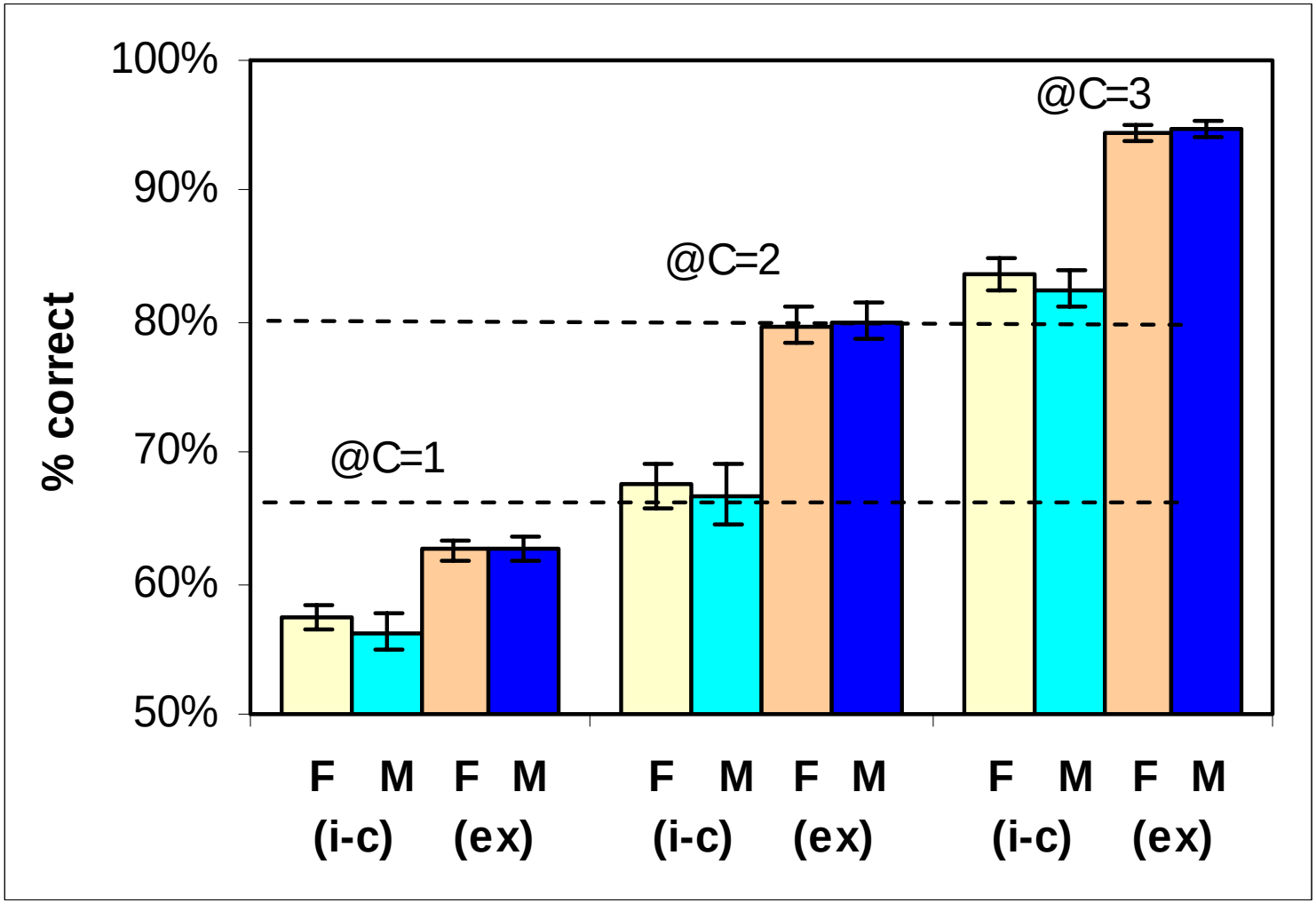
you gain:

EITHER if you find justifications for high confidence

OR if you see justifications for reservation.



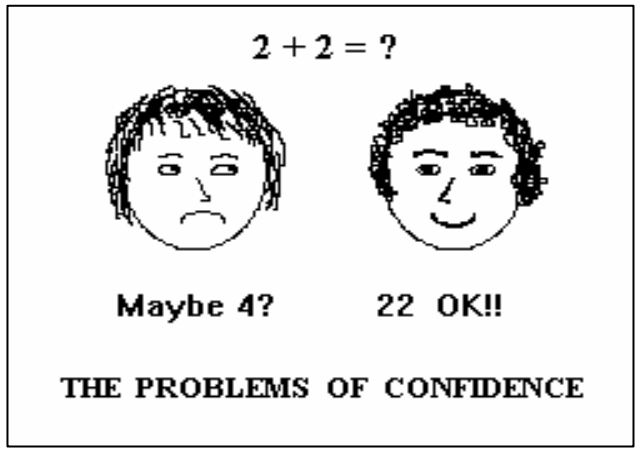
How well do students discriminate confidence?



Mean +/- 95% confidence limits, 331 students

Personality, gender issues: real or imagined?

Does confidence-based marking favour certain personality types?



- ***Both underconfidence and overconfidence are undesirable***
- ***'Correct' calibration is well defined, desirable and achievable***
- ***No significant gender differences are evident (at least after practice)***
- ***Students with confidence problems: this is the way to deal with it!***
- ***In exams, we can adjust to compensate for poor calibration, so students still benefit from distinguishing more/less reliable answers***

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%)

Reliability and Validity of Confidence-based exam marks

Exam marks are determined by:

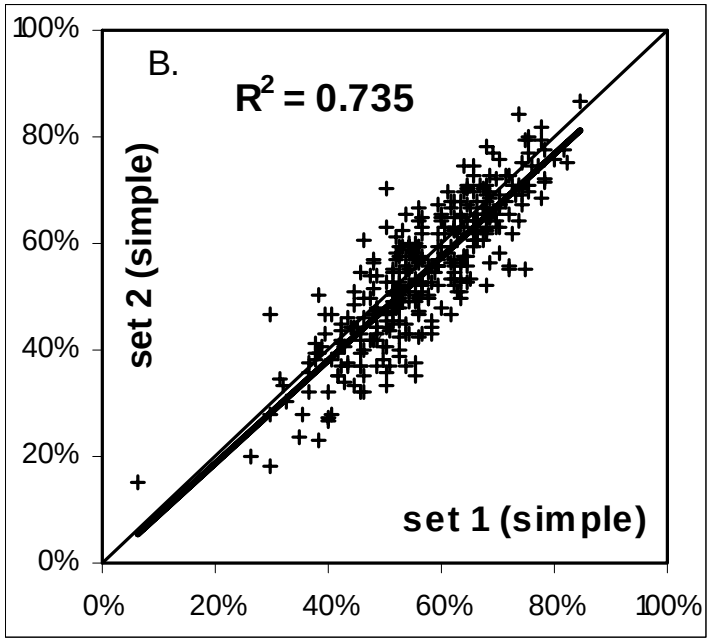
1. the student's knowledge and skills in the subject area
2. the level of difficulty of the questions
3. chance factors - how questions relate to details of the student's knowledge and how uncertainties resolve (luck)

(1) = "signal" (its measurement is the object of the exam)

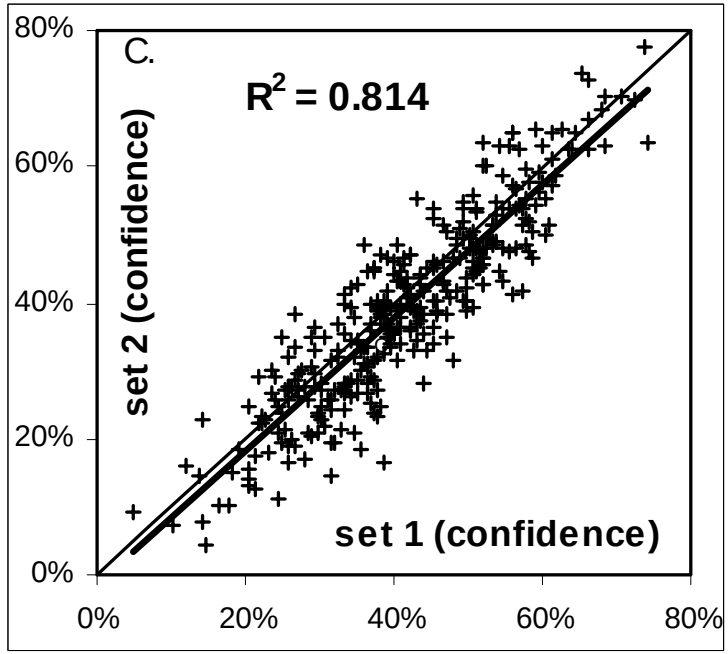
(3) = "noise" (random factors obscuring the "signal")

Confidence-based marks improve the "signal-to-noise ratio"

A simple & convincing test of this is to compare marks on one set of questions with marks for the same student on a different set (e.g. odd & even Q nos.). High correlation means the data are measuring something about the student, not just "noise".



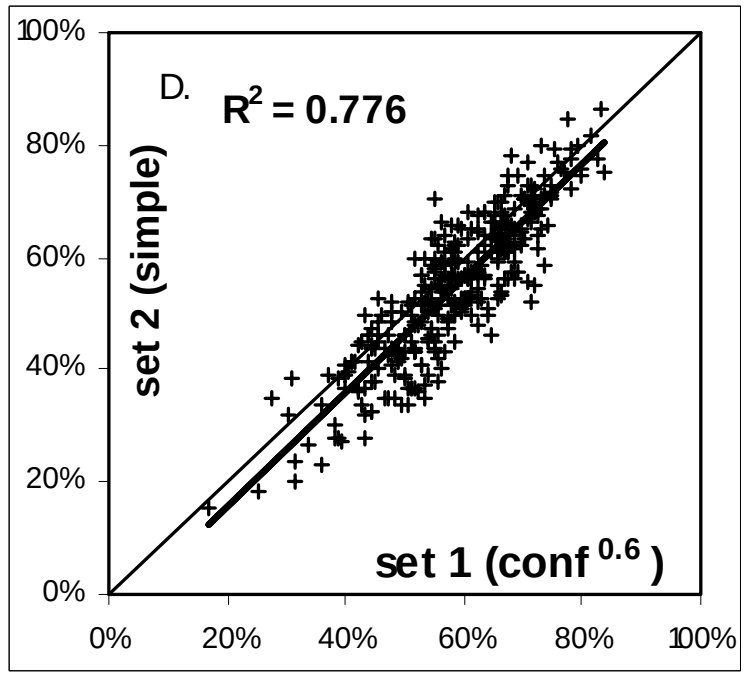
Marks scaled:
0%=chance
100%=max

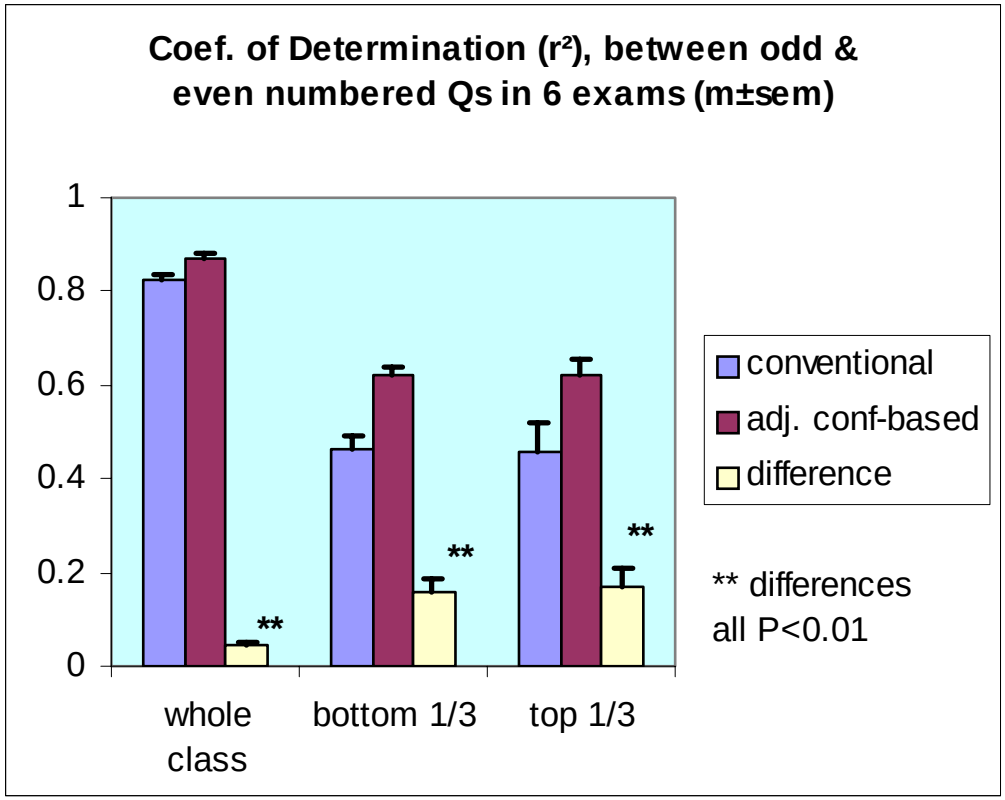


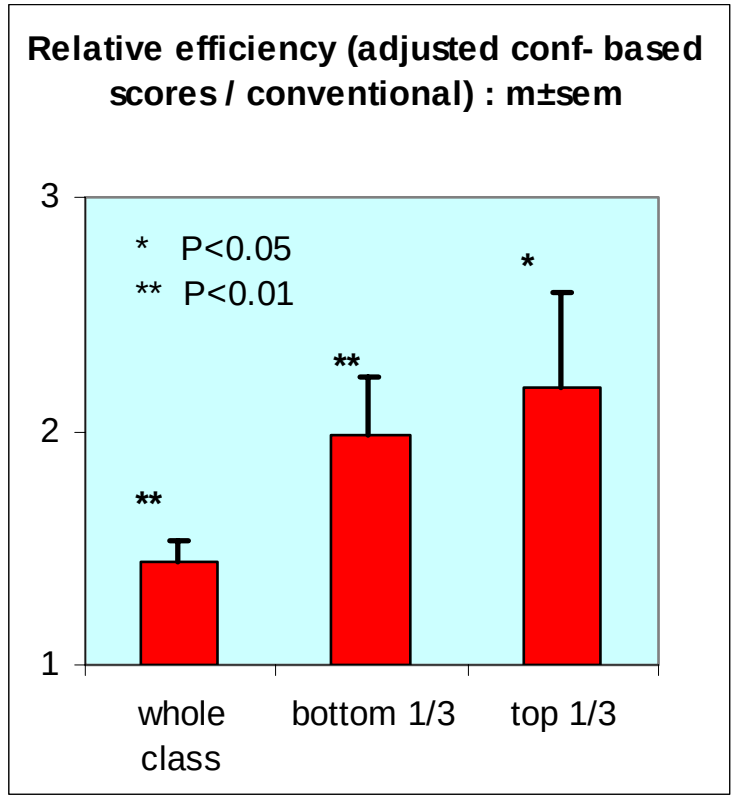
The correlation, across students, between scores on one set of questions and another is higher for CBM than for simple scores.

But perhaps they are just measuring ability to handle confidence ?

No. CBM scores are better than simple scores at predicting even the simple scores (ignoring confidence) on a different set of questions. This can only be because CBM is statistically a more efficient measure of knowledge.

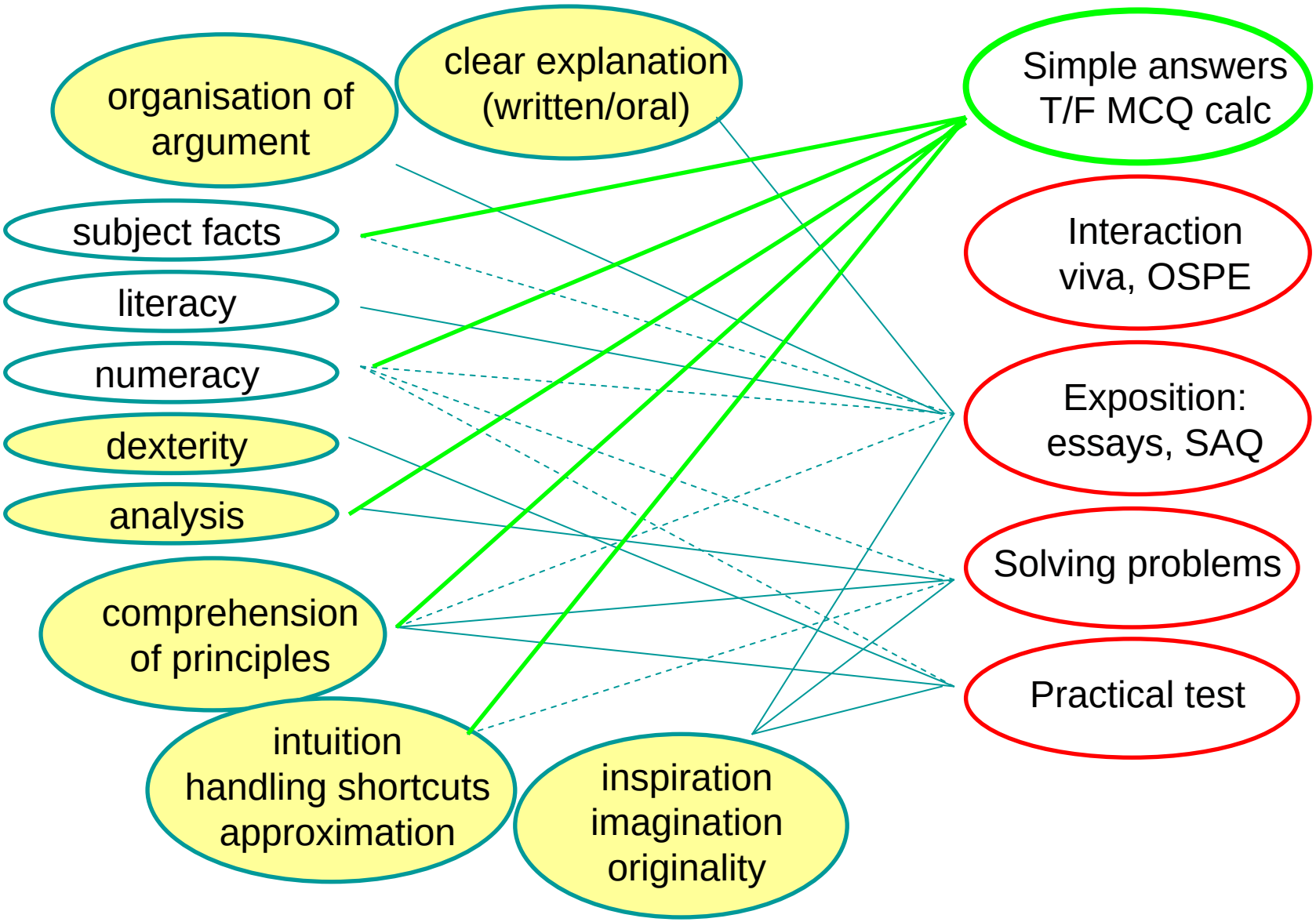






Improvements in reliability and efficiency, comparing CBM to conventional scores, in 6 medical student exams (each 250-300 T/F Qs, >300 students).

What are priorities for teachers' time ?



Knowledge depends on degree of belief, or confidence:

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

decreasing confidence in what is true, increasing confidence in what is false

Knowledge is justified true belief

In teaching we need to emphasise the importance of justification.

In assessment we need to measure degrees of belief

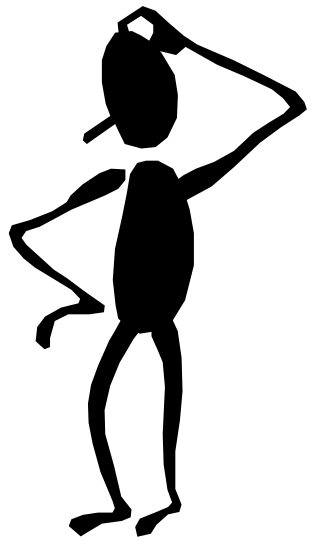
Confidence or belief (like knowledge) is a characteristic of the brain

- neither more nor less subjective than choice of an answer

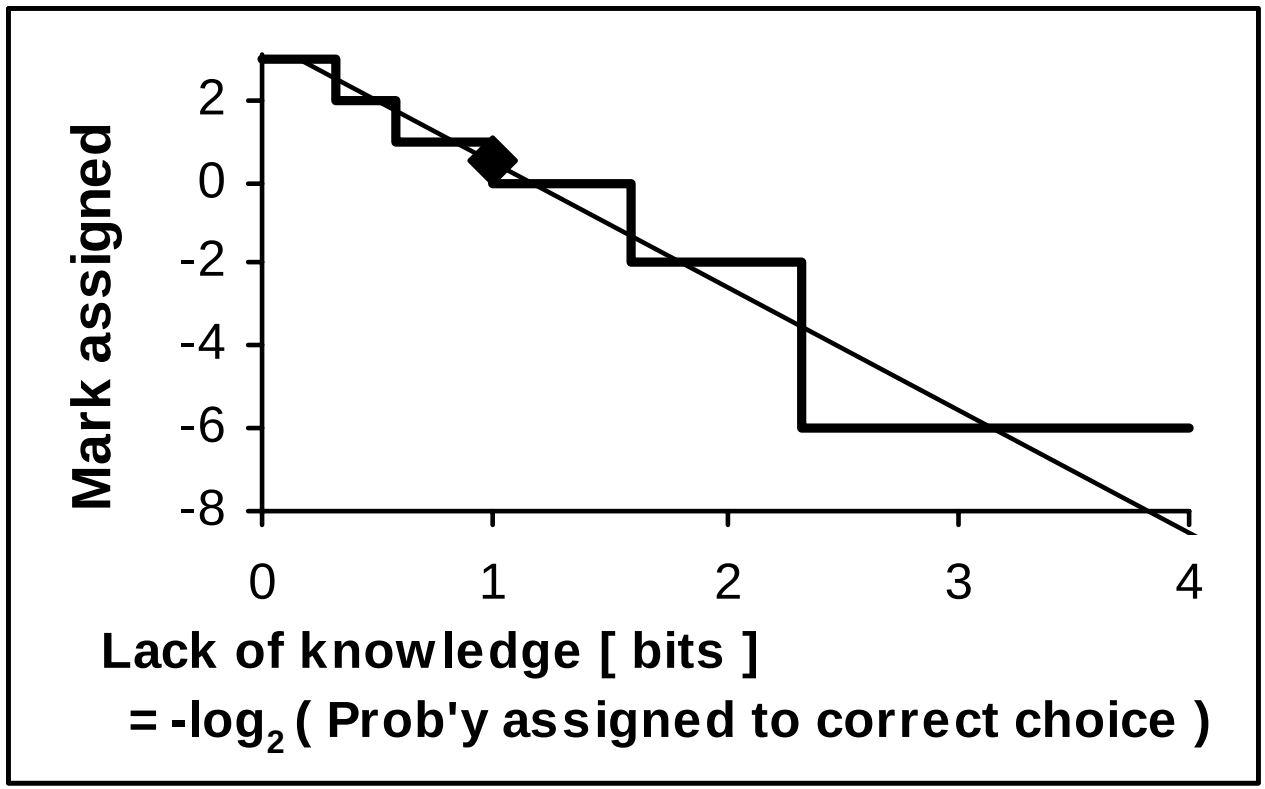
- it simply needs a motivating mark scheme to elicit it

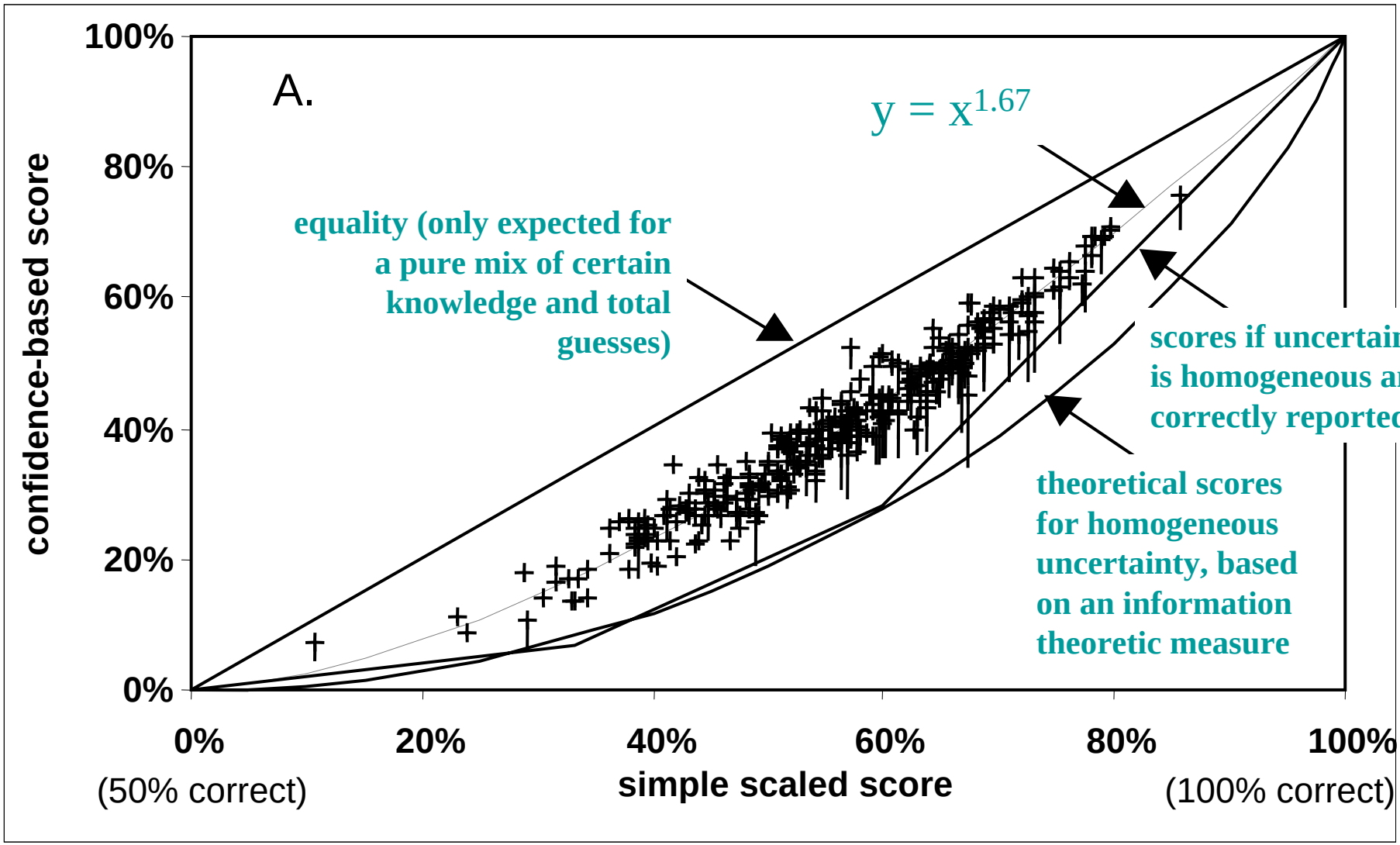
We fail if we mark a lucky guess as if it were knowledge.

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



Marks as a function of lack of knowledge defined by information theory, for T/F answers




A.
$y = x^{1.67}$
equality (only expected for a pure mix of certain knowledge and total guesses)
scores if uncertainty is homogeneous and correctly reported
theoretical scores for homogeneous uncertainty, based on an information theoretic measure
confidence-based score
100%
80%
60%
40%
20%
0%
0%
20%
40%
60%
80%
100%
(50% correct)
simple scaled score
(100% correct)

Principles that students seem readily to understand :-

- both under- and over- confidence are impediments to learning
- confident errors are far worse than acknowledged ignorance and are a wake-up call (-6!) to pay attention to explanations
- expressing uncertainty when you are uncertain is a good thing
- thinking about the basis and reliability of answers can help tie bits of knowledge together (to form “understanding”)
- checking an answer and rereading the question are worthwhile
- sound confidence judgement is a valued intellectual skill in every context, and one they can improve