Enhancing Learning and Teaching with Technology - What the research says
Ed. Rosemary Luckin UCL IOE Press, 2018

SECTION 1: What the research says about how and why learning happens.

Chapter 1.2: The value of self-tests and the acknowledgement of uncertainty

Tony Gardner-Medwin, Div. of Bioscience, University College London

Introduction

Student testing can be controversial. Why should students, especially children, repeatedly be given stressful exercises designed to rate them relative to their peers, often on measures that have limited relevance to the real world or limited value for stimulating their individual strengths and interests? It is important therefore to make it clear that the issue I examine here is not this use of formal assessment in education. Testing is not just assessment. I'm looking at the value to students of challenging themselves by doing self-tests for their own benefit, in private or with friends: learning from mistakes, discovering how arguments inter-relate, gaining immediate feedback about misconceptions and interconnections, and above all learning to think positively without the fear of humiliation that can arise from teacher interactions or exams.

Learning and knowing what we know

How do we learn things? As a teacher and neuroscientist, I, of course, know that there are many answers to this question to take account for example of different contexts and different levels. But the question is always worth bearing in mind. I learned that 7x8=56 at the age of around 8. Maybe I learned it much as a parrot could learn it - as an English phrase that just tripped off the tongue: "*rote learning*". Now, more than 60 years later, it doesn't seem to trip quite so reliably: 48? 54? 56? However, with good teaching, I learned the importance of getting such things right and how to get them right even when memory recall of the fact is unreliable. I learned that there were ways to check against things I had learned differently: my understanding of what the question means and how it relates to other memory traces: 7x7=49, so with an extra 7, this is 56. This strategy is something beyond even the smartest parrot. Confucius is credited as having said: "*Learning without thought is labour lost*". I'm not sure I go so far as this because simple associative or rote learning can be jolly useful. But in education, we need to stimulate thinking to promote deeper learning and understanding.

Challenge, self-testing and mock combat

What do we learn from examples outside the classroom? Ask any musician or sports-person how they learn and they will say the key is practice. Watch a young leopard learning to stalk prey by pouncing on its mother and you will see the same thing. These examples combine challenge, enjoyment and sometimes supreme achievement. Practice is a context where you make mistakes with impunity and without humiliation, where you can think about what you get right and wrong, and choose new situations and tactics that challenge you near your limits. That is how you expand your areas of competence and confidence. Challenge, self-testing and mock combat constitute the evolutionary origin of games, to which we owe so much in childhood and from which we can learn much in adulthood. In all our skills we each have areas of competence and inadequacy. We need good judgement about where these limits are, to work to improve them. Caution is needed when we are unsure of success, and boldness when confidence is warranted. Teachers can guide the learning process through inspiration and example or explanation, but much of the constructive work in learning is done away from public view or high-stakes competition. It involves thinking and perhaps dreaming about new challenges and things that went right and wrong. Somehow we - or at least the lucky amongst us - have been programmed and brought up to enjoy such self-testing wherever it can help.

Maths and 'Maths Anxiety'

I once had to run maths and physics classes for 1st year medical and physiology students at University College London. Though highly selected, these students have a huge range of ability in those subjects. They were not treated as priority subjects on the medical timetable: I had just two lectures, despite the fact that lack of numeracy and physical insight can have dire consequences in medicine and can be a huge handicap in science. So how can one try to deal with the all too prevalent syndrome "*I'll never be any good at maths*"? The fact is, everyone can do some maths reliably and everyone can get out of their depth at some point. My approach was to use self-tests combined with what I called "Certainty-Based Marking" (CBM) to help students establish their own boundaries. Trainee doctors must know what jobs they can do reliably and when they need help or extra time. Students were required to repeat self-tests (with randomised question selections and randomised data) as often as they wanted, indicating with each answer how sure they were that it was correct. They received instant feedback with penalties if they expressed confidence in a wrong answer. Submission of results was optional, except that at some point they had to submit at least an 80% grade on each test. This scheme was popular with students: good students passed quickly in a single trial, while weak students made as many as 6 attempts and even continued after gaining 80% because they saw that the issues being tested were issues that they really needed to understand. Even good students learned from the process since with over-confidence they often made snap decisions without thinking - eliciting a clear wake-up penalty. My approach to those who felt maths was not their thing was to encourage them to find the areas where they were genuinely competent and to extend these by identifying and working at the boundaries.

How sure are you?	C=1 (unsure)	C=2 (middling)	C=3 (pretty sure)	No Reply
Mark if correct:	1	2	3	(0)
Mark if wrong :	0	-2	-6	(0)
Chance of error?	More than 1 in 3 (>33%)	Around 1 in 4 (25%)	Less than 1 in 5 (<20%)	-

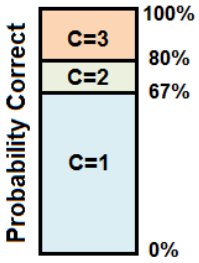

Figure. 1.2.1 A. The mark scheme used for Certainty-Based Marking. B. The rational basis for choosing a certainty level in (A), given one's judgement of the probability that an answer is correct, in order to expect (on average, with similar judgements) the highest marks.

Certainty-Based Marking

Certainty-Based Marking (CBM) in one form or another has a long history, back as far as the 1930s [Ahlgren, 1969, Hassmen & Hunt, 1994]. It is only with computers that it has become really easy to implement. But surprisingly, it is still not widely used. Post-instruction tests where students rate confidence in individual answers have been shown to enhance long-term retention. Unaware of this literature, but certain of the importance to a student of knowing where their knowledge is and is not reliable, I set up my own CBM scheme [Gardner-Medwin, 1995; http://ucl.ac.uk/lapt; http://www.tmedwin.net/cbm] (Figure. 1.2.1). It is simpler and more intuitive than earlier schemes, and properly motivates students always to be honest. It asks for a rating of certainty for each answer on one of 3 levels: C=1, C=2 and C=3. If the answer is correct they will receive 1, 2, or 3 marks (or in the USA 'points'). If the answer is wrong there is no penalty so long as uncertainty (C=1) was acknowledged, whereas with C=2 and C=3 there are penalties of -2 and -6 marks for wrong answers. With this scheme, it is obviously best to click C=3 if sure and C=1 if very unsure, with no option to 'game' the system differently to improve one's score. A rational threshold (where the blue and red lines cross in Figure. 1.2.2) is around 75% chance of being correct, and if students are unsure whether

to judge above or below this, they choose the middle option C=2. This requires thinking about reservations and justifications for an answer.

In a survey following extensive use of CBM in self-tests including core medical topics [Issroff & Gardner-Medwin, 1998] students supported its use strongly, with a majority saying they thought about confidence for most answers and sometimes revised their answers as a result. A later survey (2005, unpublished) supported continuing use of CBM in year 1 and 2 medical exams and agreed that negative marks were appropriate if answers were confident and wrong, as well as reduced marks for correct answers that were unconfident. Both at UCL and in many previous studies, CBM has been shown to improve reliability in exam grades [Ahlgren, 1969, Gardner-Medwin, & Gahan, 2003, Salehi & Sadighi, 2015]. CBM is not widely used in exams at present, perhaps due to its somehat increased complexity and the need - in the interests of fairness - to ensure that students have had prior experience with CBM. Within self-tests, CBM is more unequivocally beneficial - incentivising careful thinking and enhancing feedback. The sting of a penalty for confidence in a supposedly wrong answer helps teachers too: comments and discussion flow freely when students feel the issue unclear or contentious, so there is opportunity to clarify misconceptions and improve questions and explanations.

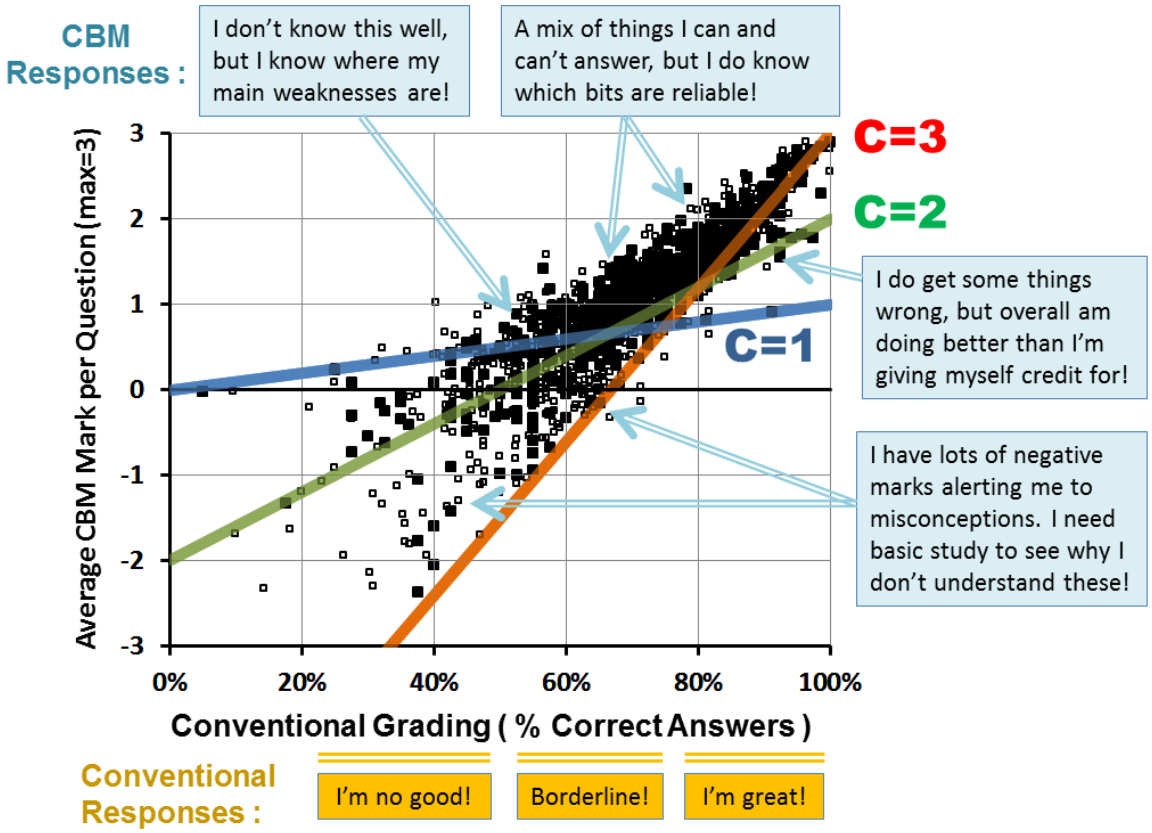

Figure. 1.2.2. Data from 1500 CBM self-test sessions taken voluntarily online [http://www.ucl.ac.uk/lapt/?bmat] as practice for the Biomedical Admissions Test [http://bmat.org.uk] used by several UK medical schools. Large symbols: students completed all questions (40-69 per test). Typical responses are shown for just conventional grades (orange) and for CBM grades (blue). Coloured lines show CBM marks that would have been awarded if students always used the same certainty level. About 60% of grades are above these lines, indicating successful discrimination. This is reflected as a bonus in feedback to the students, added to their conventional grade.

CBM provides a new quality of feedback to students through self-testing. Figure 1.2.2 shows data from subjects previously unfamiliar with CBM, practising online for a prospective application to medical school. Feedback with just conventional grades on such tests can lead to crude and often

dispiriting student responses (orange). Adding the information from CBM (the vertical axis) allows much clearer categorisation of a student's status and issues (blue). These are difficult self-tests, and about 10% of the subjects obtained negative overall CBM marks. Since it is obvious that they would have done better by acknowledging uncertainty (C=1) all the time, the feedback provides a wake-up call that, even with quite good conventional grades over 60%, they may need to address issues where they don't realise they are weak. Such negative overall CBM scores are of course much more rare in final exams, since misconceptions (when defined as confident wrong ideas) have, for most students, largely been dispelled. CBM tests at the end of a course may show a lack of knowledge, but typically indicate good discrimination between what is and is not known. Gender biases sometimes observed with conventional (especially negative) marking - perhaps due to females being more inclined to disadvantage themselves by omitting uncertain answers - seem to be absent or much reduced with CBM [Gardner-Medwin, 1995, Hassmen & Hunt, 1994, Salehi & Sadighi, 2015].

Voluntary Self-Tests in a Medical Course

At Imperial College Medical School, with Prof. Nancy Curtin and colleagues, we have kept track since 2008 of the relationship between voluntary use of CBM self-tests in the first year course (October - June) and pass/fail rates in formative (January) and summative (June) exams [Curtin & Gardner-Medwin, 2013]. The self-tests were mostly drafted by students in previous years, with vetting by staff. The statistics over 7 years (averaged in Figure. 1.2.3) are striking. Failure rates for students who used relevant self-tests before the exams were, in both contexts, consistently and markedly lower than for students who had not used self-tests. After the first year, students were each year (at the end of Term 1 and before the first exam) shown such data from previous years in lectures; but the discrepancy nevertheless persisted.

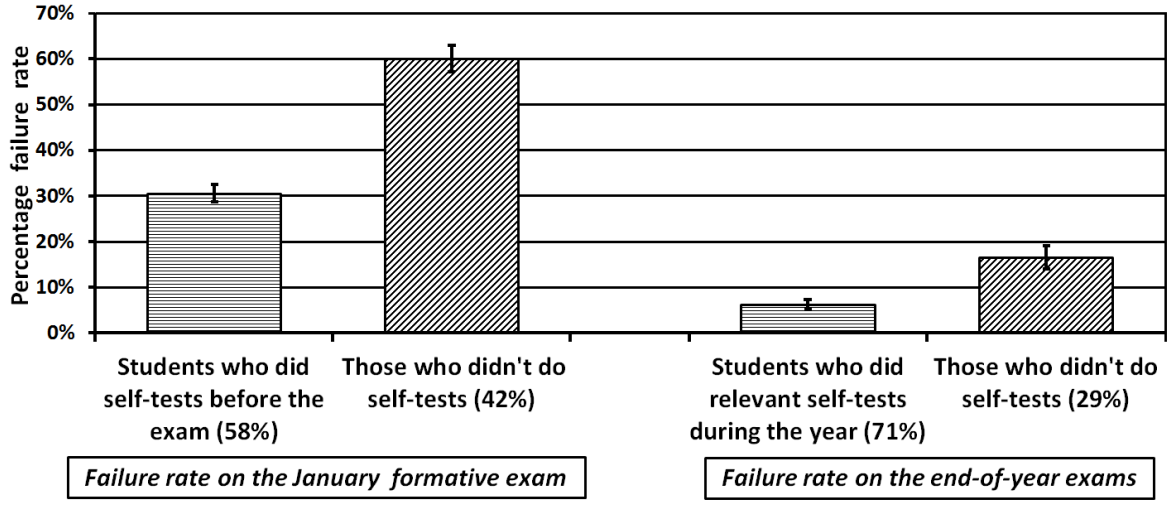

Figure. 1.2.3 Failure rates in exams for students who did and did not choose to attempt CBM self-tests in year 1 of a London medical course. The exams did not employ CBM. Averages (± 1 ± s.e.m.) over 7 years 2008-2015. Student numbers 270-330 per annum.

We do not suggest that the students who tried self-tests necessarily did better in exams because of the self-tests. We certainly hope these may have contributed, but good students are doubtless more likely both to use available learning tools and to pass exams. Around 30% of the students who failed the formative exam without trying self-tests did start using them later in the year. But it is worrying that many students who were destined to fail these exams did not have the insight to foresee the problem, and at least try something that the data suggested might help. This lack of insight is exactly where self-tests, especially with CBM, should help: identification of weak areas through practice, challenge and mistakes. Medical students are smart and have all done well at school. Perhaps this is part of the problem: some of them need to learn that in a wider world there are

always limits to what you understand, and without self-testing, it is difficult to realise how uncomfortably close these limits may be.

Conclusion

To quote Confucius again, "*Real knowledge is to know the extent of one's ignorance*". The most valuable element of self-tests and CBM is perhaps to reward acknowledgement of uncertainty as something more valuable than self-confidence. Teachers often have trouble getting students to acknowledge uncertainty. In the privacy of a self-test and with the unconventional structure of certainty-based marking it is possible to reap some of the benefits of Confucius's insight.

References

Ahlgren (1969) Reliability, Predictive Validity, and Personality Bias of Confidence Weighted Scores. Proc. Amer. Educ. Res. Assoc. Available at http://assessmentpriorart.org/tag/confidence-based-marking/

Curtin N & Gardner-Medwin AR (2013) Self-tests with Certainty-Based-Marking in Early Years of Medical Course. *Int. Union Physiological Sciences 2013, Birmingham* POSTER available at http://www.tmedwin.net/~ucgbarg

Gardner-Medwin AR (1995) Confidence Assessment in the Teaching of Basic Science. Research in Learning Technology 3:80-85

Gardner-Medwin AR & Gahan M (2003). Formative and Summative Confidence-Based Assessment. IN: Proceedings of the 7th CAA Conference, Loughborough. Available at: https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/1910

Hassmen P, Hunt DP (1994) *Human self-assessment in multiple-choice testing.* J Educational Measurement 31, 149-160.

Issroff K & Gardner-Medwin AR (1998) Evaluation of Confidence Assessment within Optional Coursework. , In Innovation in the evaluation of Learning Technology. Ed. M.Oliver, Univ. N. London Press, pp169-179.

Salehi M & Sadighi F (2015) Comparing Confidence-based and Conventional Scoring Methods: The Case of an English Grammar Class. *The Journal of Teaching Language Skills* 6:123-152