

Perspective on Certainty-Based Marking

An Interview with Tony Gardner-Medwin

Reid Cornwell and Tony Gardner-Medwin

Note: This article was originally published as a Multimedia file in *Innovate* (<http://www.innovateonline.info/>): Cornwell, R. & T. Gardner-Medwin. 2008. Perspective on certainty-based marking. *Innovate* 4 (3). <http://www.innovateonline.info/index.php?view=article&id=552> (accessed February 5, 2008). The transcript is reproduced here [lightly edited] with permission of the publisher, [The Fischler School of Education and Human Services at Nova Southeastern University](#)

RC: Welcome to Perspectives. I am Reid Cornwell, your host. My guest is Dr Tony Gardner-Medwin, Professor emeritus of Physiology at University College London. In this edition, we will be discussing enhancing Learning, Teaching and Assessment through Certainty-Based Marking. Tony, thank you for speaking with me today.

TGM: I am very pleased to have the opportunity to speak with you.

RC: Could you explain how you are using the term marking in this context?



Reid Cornwell



TGM: It's about assessment; how we mark students' work. Actually assessment is something that teachers don't usually regard as the favourite part of their job! But it's absolutely crucial to learning for students. Firstly of course, assessment drives learning in the sense that students are motivated always to get good grades. Then the ways that we assess students are very strongly influential in how the students study and think. Thirdly, students need to learn to self-assess, to judge their own skills and their own abilities, from thinking about their own work. And lastly they need to practice. Self-assessment gives practice at very low

cost to teachers and with little risk of humiliation to the students. Education is really about learning how to think and how to relate things. It shouldn't just focus on students learning to repeat things. And Certainty Based Marking is aimed at trying to get students to think more when they're answering questions, which is part obviously - part of any kind of course.

RC: Tony, I think of marking as grading, and assessment as what the instructor does with the information derived from marking. Do you recognise that difference?

TGM: Well, assessment is a very interesting word, because I think its origin is actually sitting beside somebody and interacting with them in a way that helps them to work better. In universities and schools, I think we've got rather away from that concept of assessment. Assessment of a student, of course, should be a rather wide-ranging thing and it should include aspects of their ability which are very diverse. For example, I feel that a transcript of a student's performance on a course should

ideally include assessments of how good they are at expressing their knowledge, how good the breadth of their knowledge is, the reliability of it, their insight into problems, and their ability to be creative about forming solutions to problems. So, I believe that overall assessment is a very broad ranging thing and shouldn't be expressed in a single mark. But if you look at the motivation for using CBM, it's to try to improve the student's attitude to individual little nuggets of what they are doing in a course. They're confronted with a question, and we're trying to improve their judgment about how best to get at the answer to that question, how best to justify the answer and how to judge how reliable the answer to that question is. And, on individual points where they're being asked a question - especially when it's a computer-based question - assessment is bound to be indicated in a rather simple form, usually just a single number. And this is where the Certainty-Based Marking does better than just making that number one or zero depending on whether they're right or wrong - it actually takes account of how they judge the reliability of their answer. So, marking is simply the generation of a number for a very small element of the overall assessment. This is my view of the relation between the two words.

RC: What are the mechanics of this system?

TGM: Essentially it's very simple. The students, whenever they're asked a question, will be marked right or wrong. They also say how certain they are that they're getting it right, on a scale of one, two or three. If they do get it right, they'll get one, two or three marks. So, obviously it's in their interests if they're sure of the answer, to say that they are sure - then they get three marks if they are right. But if in that situation they're wrong, they get a severe penalty: a double penalty, minus six, twice what they stood to gain if they'd been right.

Finish Sect. 4 (5 question/s)

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)

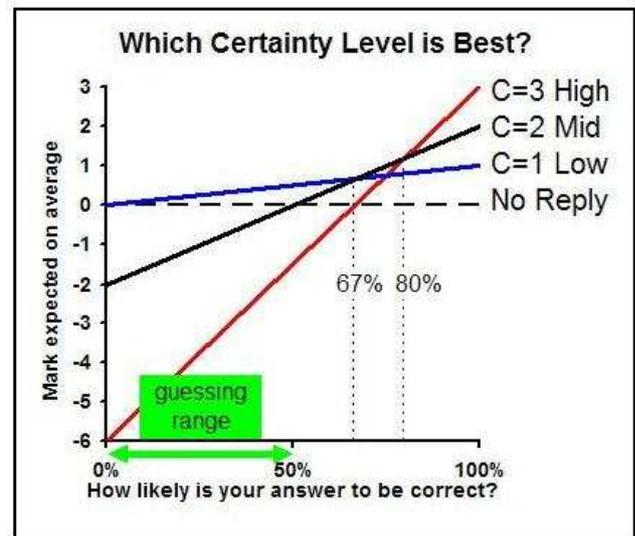
So you have to be careful, and think carefully, before you say that you are sure that you are getting an answer right. If you opt to say that you are actually uncertain of the answer, then you'll only get one point if you are right, but you get no penalty if you are wrong. So, it's actually to the student's advantage to think about the subject properly, even if they come up with reasons for being unsure of their ground, whether or not they stick with an initial answer. Whether they end up thinking that they're sure or unsure, the result of their thinking will (on average) benefit them.

RC: So, certainty in this context is also analogous to confidence in what one knows? Is that a correct assessment?

TGM: Yes. Confidence, of course, is something that people can have as a general thing. People may be self-confident in their personality or may be very unconfident in their personality. But we're really talking about confidence about a very specific thing. Somebody who is not a very confident person may indeed be certain that they're getting a particular answer right. So when I talk about Certainty-Based Marking, it's based on how certain you are of that particular answer, not of how confident you are in general.

RC: By addressing certainty, it appears that you also address guessing.

TGM: Well, the guessing issue is very interesting, and actually I regard the way conventional marking treats guessing as completely indefensible in one way, because when a student makes a guess at an answer, and it's a lucky guess and they get it right, then a conventional scheme treats this as exactly the same as if the student knew the answer, which they did not - they had a lucky guess. So, I think, in principle you could sue an exam board for regarding those two things as equivalent. [I think this because] in fact there is a proper way to deal with the issue of guessing - which is to give more weight to things that the student knows than to things the student gets by guessing right. The only way to deal with this really is Certainty-Based Marking [in one form or another]. You don't want to discourage students from guessing, or at least making a partial guess. What you want to do is reward their acknowledgement that they are uncertain. And this is exactly what CBM does. If you are uncertain, then you will expect to do better on average by acknowledging your uncertainty [as shown in the graph]. At the same time CBM decreases the relative weighting for uncertain answers. So I think guessing is actually a fundamental issue that should not be dodged, and the way to deal with it in assessment is by CBM, Certainty-Based Marking.



RC: Why should teachers and learners use CBM?

TGM: Well, I think everybody agrees that it's important that you should think about how reliable your ideas are. It's obviously fundamental to doing well in studying and doing well in applying your knowledge. Secondly is the issue of trying to encourage understanding of the issues, not just reacting immediately to a question. We find with a lot of our students - medical students are highly selected - they are very used to being able to pass exams based on just the first thing that comes into their head, and they don't necessarily think really beyond that sort of level in response to a lot of questions; they just answer it straight away and don't think more. And really what we want to do is to try to raise the stakes on answering questions, so that they need to think about whether they understand what's going on, in order to be able to tell whether they can justify it or whether they can see that there's a problem - and that maybe there is good reason to be uncertain.

We want to encourage students to think laterally: in other words, to try to think whether there are issues that are perhaps not the most obvious, that also bear on the question that they are being asked. And to try to link different pieces of information that they have knowledge about together in a sort of network of knowledge that is really what people who are proficient in a subject have developed. When they're asked a question, they don't just answer it on the basis of one thing, but they can see why the answer *has* to be right because of all sorts of other pieces of knowledge that bear on that, and that relate to it. The next thing here is that you really don't know something unless you're prepared to take a risk on the answer that you've given. If you just get the thing right because you made a lucky guess, that really is not knowledge. You really should be challenged, as to whether you're prepared to bet on the answer, if you like. This is really what life is all about.

There's an interesting issue that comes up about people who are actually not very self-confident about their knowledge in the subject, that they can often lose out in teaching and learning situations because through their lack of confidence, self-confidence, they often won't even be prepared to put their hand up, or sort of volunteer ideas in a class situation. But they may actually be

much better able to deal with the subject than they're giving themselves credit for. And through the Certainty-Based Marking they actually can learn that even if they start out by putting $C=1$ for a lot of their answers, that they will learn that this is actually not expressing their true knowledge, that they actually would get a better mark if they raised their certainty and if they were less diffident; and this helps their self-confidence.

Next, the students seem immediately themselves to appreciate that it's a more fair way of assessing whether they actually know the answers to questions, - that a thoughtful answer that they can justify and see *has* to be right deserves more credit than something they think they've remembered right, but they're not really very sure - but they'll put it down all the same! They know that the first kind of situation deserves more marks than the second kind of situation, and the CBM scheme actually relates very directly to the way the students understand the process of assessment and self-assessment.

There's a very important point that arises when students make confident errors. When they get a minus six on the CBM scheme, this means that they are pretty sure that they're getting their answer right and that they've in fact been marked wrong. Now there are various things that may be true here; the first and most common is that the student has a misconception about the subject, and really the minus six that they get in this kind of situation is a wake-up call, it helps to stimulate them to actually think more carefully about the information - the explanations that may be given in a learning situation in response to their answer. So it helps them to reflect about whether they really have this piece of the subject straight - maybe they should go back and read the textbook about this and see why they were so sure about something that was wrong. That's the most common situation. Of course, also sometimes it's true that that question has been rather poorly phrased and that it's maybe ambiguous or maybe its plain wrong. This is actually a situation of great value to the teacher because the minus six in that kind of situation often leads to a sort of outrage on the part of the student, that they've been penalized for misunderstanding a question that wasn't even clear in the first place. And what we find is that the students are very inclined to enter a comment there, and explain why they did have this reaction to the question; and this helps the teachers to improve the quality of exercises that they're giving to students in a formative self-assessment situation.

Lastly is just a general educational point, that if you're going to be able to study efficiently, then you must be constantly questioning what you know, and thinking about the justification for what you know. When we read a paragraph explaining something, most of us in universities, we learn to constantly ask ourselves whether we understand what's gone before and whether we could make predictions from that and whether we could actually generate the ideas that are following on later on. The idea that you should, along with any idea that you generate, generate also a judgment about how reliable it is, is very fundamental to the whole process of learning and studying.

RC: Tony, you wrote on your help page, 'knowledge is knowing what you know and what you don't know'. This is similar to Confucius. Does this reflect his influence?

TGM: Oh gracious! Well I think what Confucius says is obviously right. It's so deeply ingrained in any kind of academic interaction in a university context or a scientific context - that when you believe something you must also be able to justify it and use that justification to persuade somebody else it's true. And it's a terrible academic sin to claim something is true if you can't justify it. So, what Confucius said, I think, is obviously very much right. There's actually another thing that Confucius said that appeals to me in the educational context, which is that learning without thought is essentially a waste of time, and this is what I was saying earlier on, that indeed if a student learns something but they don't understand and they don't think about what it means, then it is a completely unproductive kind of learning, and it shouldn't be encouraged. So this is one of the things that we're trying to prevent students doing - simply learning to repeat things.

RC: You have written that CBM tests knowledge not facts. What is your definition of knowledge?

TGM: Ok. Well, there's of course a long philosophical literature on what the definition of knowledge is, and - really in everybody's book - it requires that the information be right, be correct, and that it be something that you believe - which is a matter of the probability that you assign to its being correct. So the certainty that you assign to something is very fundamentally a part of whether something is knowledge. And the third thing is justification. You must be somehow able to justify the answer that you are giving. And the important thing that CBM does is that it requires the student to think about whether they are confident in their answer or whether they're not confident in their answer. That's very fundamental; and in order to do that they really need to think about whether they *can* justify the answer. And if they can justify the answer and come to the conclusion that there's good reason to be sure it's right, then that's great; they'll get more credit as long as they *are* right. If they also, through thinking about it, they come to the conclusion that there's reasons for uncertainty about their answer, then as long as they acknowledge that uncertainty, they'll also gain by having thought through the issues.

In fact, when you're marking student material, the thing that really gets under my skin as it were, is when a student says something that is right, and you're pleased about that, but then, a little bit later, it turns out from something else they say or something that you get out of them by questioning, it turns out that the thing that they said that is right, is something that they don't actually even understand what it means. This is not knowledge, it's really just the ability to repeat facts, or repeat not even facts: it's repeating the expression - the words that would express facts if you did understand those words. It's really important that you should not regard just the repetition of sentences as being knowledge. Knowledge must be the ability to justify the meaning of what you're actually saying. So that's really what I mean by knowledge. It's the ability to justify things that you have good confidence in and that are actually correct.

RC: Tony, tell us how Certainty-Based Marking would fit into a teaching strategy.

TGM: Well, Certainty-Based Marking is really [mostly] used through computer assessment. And of course in a teaching programme the most important thing is the teachers and how they actually interact with the students. And there are many aspects of teaching and many aspects of assessment that the teachers are essential for - where they must use their human skills. The Certainty-Based Marking makes the computers much more useful and effective in providing assistance to the teachers to free their time for that sort of a thing. It also makes the students much more receptive to self-assessment on computers, and it helps to improve their study habits, I think, as a result. So really, I'm far from an advocate of computers being a cure-all for all the problems in education, but we really must use them efficiently, and I think Certainty-Based Marking helps in that process, using them as efficiently as we possibly can.

RC: What are the educational goals of this system?

TGM: The main aim of introducing Certainty-Based Marking is to try to get students to think more about the questions that they're asked. So it's really a part of the study that they do, as the prime thing; we also use it for exams for the students, but the most important aspect of Certainty-Based Marking, in my view, is for students doing revision; students thinking about questions that they find difficult. Many students do very well in exams without really thinking very deeply about the things they're being asked about. The first thing that comes into their head is often good enough; it's often right; and they don't need to think any more deeply about that. But that's not really satisfactory, and it's actually damaging to their whole attitude to learning.

RC: Tony, can you summarise the findings that support the efficacy of CBM.

TGM: The kinds of evidence that we've accumulated [Ref 1] are, firstly, informal reactions from the students and evaluation data from surveys of the students. Then there's of course the results from analysis of the tests they've done and exams they've done. We haven't tried to do any sort of formal comparison with students who haven't used CBM, because it's very difficult to manage in a University.

The first thing that was obvious was that the students immediately see the benefits in CBM, and they don't seem to have any trouble understanding how to use it and the logic of it, and they actually - with practice - they get very good at using it in a nearly optimal calibrated kind of way. So that has not been any problem.

We actually have used it in exams for five years or so at UCL, and we surveyed them about their attitudes to this, and they voted really quite strongly, I forget exactly, I think it was sort of 55 to 35% or something like that [actually 52% : 30%] in favour of keeping CBM in the exams that they were having in the first two years of the medical course. In the exams the data has shown that it has been really quite successful. One of the ways of assessing exam data is what's called reliability, which is really a measure of how well the exam is measuring something about the student rather than about chance factors: how lucky the student is with particular questions. You can always increase the statistical reliability of an exam by increasing its length, using more questions, but when we compare the CBM with conventional marking we get an increase of reliability which is really quite substantial. It corresponds to an increase in a test length of about 58% I think it was on average. And it showed up - the increase of reliability - it showed up in every one of the seventeen exams that we analysed in data that is on the website [Ref 2].

In addition to this reliability business, there's a concept of validity, whether you're really measuring what you want. You might argue that you don't want your measurements to be dependent on how well students use a CBM system: what you want is to be measuring their knowledge defined in some other way. Even if you define their knowledge as the number of questions they get right in a test, regardless of their certainty about the answers, we find that the best predictor of that is actually the CBM score on a different set of questions. If you divide the exam into two sets of questions randomly, then if you want to predict the score on one set, the best thing to use is the CBM score on the other set, not the conventional mark - even if your gold standard of what knowledge is, is the conventional mark.

One of the concerns that there was at the beginning was that there might be biases, that there might be gender differences for example in the way in which students use CBM. People suggested that this might be unfair in penalizing students who weren't very confident, and they generally suggested it was going to discriminate against females. We've actually looked very carefully, because of that, at the data from our exams and from our online use - and there's absolutely no suggestion of any gender bias. Comparing ethnic groups - and we have quite a mixed group of medical students at UCL, but unfortunately not such reliable relevant data - analysis has again failed to suggest any significant differences in the relation between indicated certainty levels and the actual reliability of answers. Interestingly, we find quite large differences between the risk aversion in exams compared with the risk aversion when the students are working online: they tend to be much more cautious about indicating high confidence in exams. This shows that the data is quite sensitive to differences in behaviour, but there's no suggestion of the biases that one might be worried about.

As I said, we don't really have data that the graduates end up better if they've used CBM in tests than if they haven't, but really this would require a randomized controlled trial. With students in a medical school that would actually be very difficult, because one of the things that we always encourage in a medical course is that students should work together. This means that if you try to allocate them to groups that did CBM and didn't, this collaboration between students would rather defeat the purpose, and it would be very difficult to draw conclusions. So, I mean, I'm rather opting

out of trying to answer a question about whether there is definitely an improvement of the quality of the graduates. But I think the fact they've been exposed to a new kind of technique for trying to improve their ways of thinking about their study is really sufficient to indicate that there's benefit in doing this rather than not doing it.

Whether you're talking about kids at school or college students, and however good or bad the students may be, the questions you ask them should actually engage them with the subject, and this essentially is what makes learning fun for students - fun and interesting.

RC: Tony, thank you for sharing your innovative work and your thoughts.

TGM: Well, it's been a great pleasure. I really encourage people, if they want to find out more about Certainty-Based Marking, to try it out and look at the website [Ref 3], which will help them to do that in any kind of context that they're interested in.

1. <http://www.ucl.ac.uk/~ucgbarg/pubteach.htm>
2. http://www.ucl.ac.uk/~ucgbarg/tea/UCL06_cbm_poster.pdf
3. <http://www.ucl.ac.uk/lapt>