Technische Universiteit
Eindhoven
University of Technology

Evaluation of Certainty-Based Marking and Learning using Oncourse/Moodle

Version 1.1 of 1 December 2015

Kees Goossens

ES Reports
ISSN 1574-9517

ESR-2015-04
16 November 2015

Eindhoven University of Technology
Department of Electrical Engineering
Electronic Systems

http://www.es.ele.tue.nl/esreports
esreports@es.ele.tue.nl

Eindhoven University of Technology
Department of Electrical Engineering
Electronic Systems
PO Box 513
NL-5600 MB Eindhoven
The Netherlands

Evaluation of Certainty-Based Marking and Learning using Oncourse/Moodle

Kees Goossens

Version 1.1, dated December 1, 2015

1 Version History

- 1.1 Added minor comment on student feedback.

2 Introduction

Certainty-based learning (CBL), also known as confidence-based learning, is a learning approach [1] that not only asks students to give answers to questions (in examinations, homework, and so on), but additionally asks students to indicate how certain they are of their answer.

In traditional examinations it is not possible to distinguish between correct answers that the student was certain of and lucky guesses. However, it is clear that knowledge is not just about the right answer, but also about the certainty with which the answer is given (i.e. the *insight of the student into the validity of his or her knowledge*). Additionally, it may be argued that it is fairer to give a higher grade to a correct answer given with a higher certainty. (And, conversely, a lower grade to wrong answer given with a higher certainty.) To implement this intuition, Certainty-Based Marking (CBM) is a marking method that determines the grade based on the correctness of the answer as well as the certainty with which the answer is given.

It has been shown that confidence in answers is a better predictor for knowledge retention than the accuracy (percentage of correct questions) [2]. In other words, if as educators we aim for *long-term knowledge retention* (i.e. students should not just pass for an exam and instantly forget what they have learned), we should take confidence in answers into account in the grade.

The website [3] of A.R. Gardner-Medwin of the University College of London (UCL) contains a host of information on the topic, including on reliability of CBM vs. traditional tests. (For example, in an exam, CBM scores on odd questions are a better predictor for the CBM scores on even questions than when using traditional percentage correct.) CBL has been used for many years at UCL. In 2006 students voted 52%:30% to retain CBL [4].

CBM can be applied to homework and (intermediate) examinations, and is supported in Oncourse/Moodle quizzes.

3 How it works

Every traditional question (open, multiple choice, etc.) has a follow-up multiple-choice question on the form:

What is your certainty?
C=1 (Unsure: <67%) C=2 (Mid: >67%) C=3 (Quite sure: >80%)

In Oncourse this would look like shown in Figure 1.

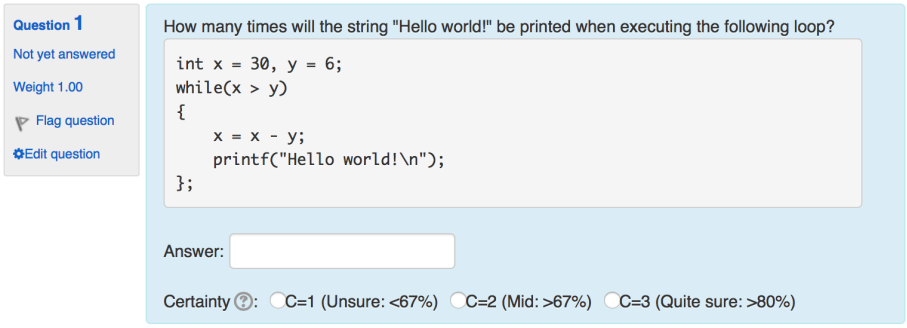

Figure 1: CBM question in Oncourse.

The grade of a question is determined by the following table:

degree of certainty	C=1 (low)	C=2 (mid)	C=3 (high)	no reply
grade if correct	1	2	3	0
grade if wrong	0	-2	-6	0

A grade therefore ranges from -6 to +3, or from delusion, misconception, ignorance, to knowledge. Right answers obtain higher grades. But importantly, *with increasing confidence wrong answers are punished more severely than right answers are rewarded.* As a result, it never makes sense to gamble or to lie about certainty. In particular, gambling or 50% probability of being correct, gives an average grade of 0.5, 0, or -3 at low, mid, and high confidence respectively. It is quite intuitive that there is no way to "game" or "play" the system: the highest grade is obtained by honestly indicating the certainty.

The magic numbers of 67% and 80% are in fact clearly defined as the intersections of lines equal average certainty, see Figure 2. To maximise your score:

- if you are correct $< 67\%$ of time you should pick C=low for all questions

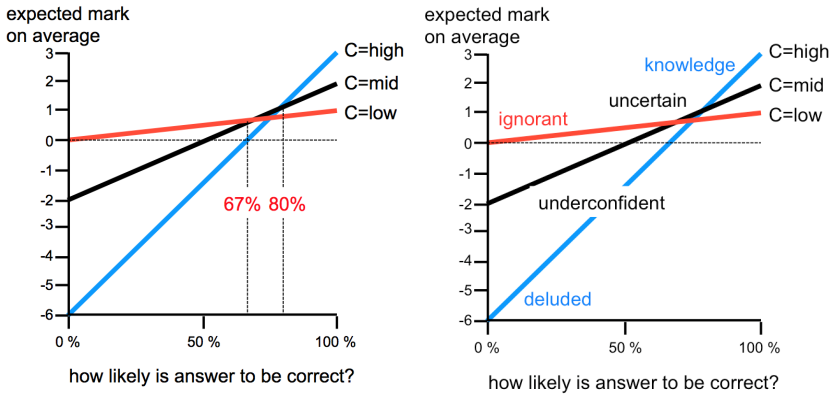

Figure 2: Certainty-based grading, reproduced from [5].

- if you are correct >= 67% of time you should pick C=mid for all questions
- if you are correct > 80% of time you should pick C=high for all questions

Interestingly, the weights -6, -2, 0, 1, 2, 3 lie linearly on a log scale, indicating the (lack of) information (knowledge, ignorance, misconception, delusion), as shown in Figure 3.

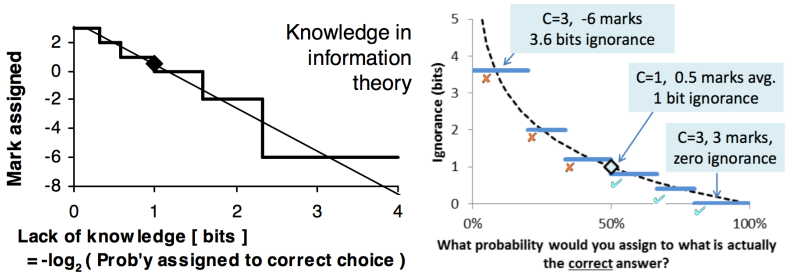

Figure 3: CBM weights, reproduced from [4] and [5].

Certainty-based grades (from -6 to +3) are useful in self learning, e.g. with Oncourse quizzes that give feedback on the student's performance. A student can do the same homework, practice exam, etc. multiple times and receive feedback on accuracy and certainty each time. Figure 4 shows the result of a real examination with a maximum of 37.00 points, of which 28.00 were obtained by the student. Note the feedback on how the student rates his knowledge in the break-down by certainty. Even though this student had only 40.5% of the questions correct (Accuracy), he clearly knew when he was correct (6 questions), and where he wasn't sure. As we shall see in the next Section, this resulted in a higher grade than with traditional accuracy-only grading.

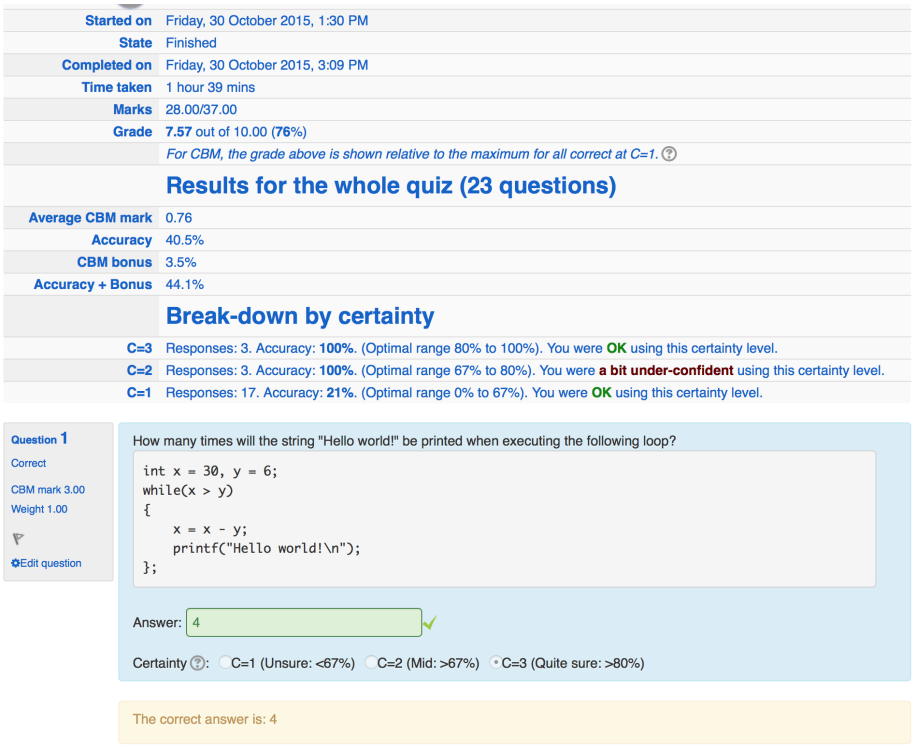

Figure 4: CBM feedback in Oncourse.

4 Converting a CBM Grade to a 1-10 Grade

So far, CBM and CBL are quite intuitive and easy to explain and justify to students. The story becomes harder to explain when it comes to converting the CBM grade (ranging from -6 to +3) to our 1 to 10 grading range. This cannot be done linearly, and there are two schemes to do this [1]. For the scheme we use, in essence, the grade is the traditional percentage of correct questions plus a bonus for correctly-applied certainty. *The bonus reflects the insight of the student in his or her knowledge.* The bonus is positive or negative, proportional to the amount the average CBM grade is above (or below) the average that would be obtained if the student had used the same optimal confidence level for all his or her answers [5]. The grade (known as CB accuracy) is computed as 10% $\times$ (actual average CBM grade $-$ average obtainable with uniform certainty). This is illustrated in Figure 5. CB accuracy, which we will call converted CB grade,

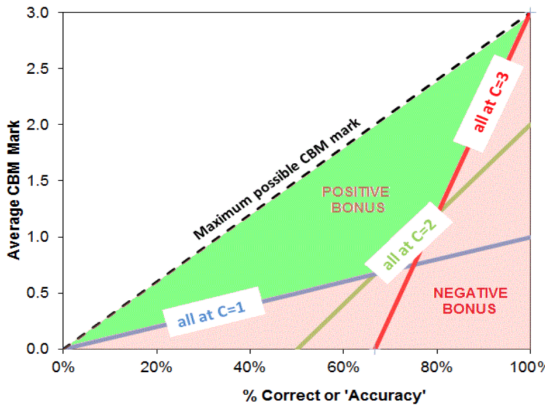

Figure 5: CBM accuracy and bonus [5].

is computed by the following equations:

```
accuracy = number of correct answers / number of questions
cbm average = CBM score of answers / number of questions
bonus = 0.1 * (cbm average - max (accuracy, -2*4accuracy, -6+9*accuracy))
converted cb grade (0-10) = accuracy + bonus
converted cb grade (1-10) = 1+ (converted cb grade)/10*9
```

(Minor modifications to accuracy and CBM average take into account questions with different weights and partially correct questions.)

Consider the student of Figure 4 as an example. The traditional accuracy, i.e. number of correct questions, is shown as 40.5%. Because the student correctly indicated higher certainty on correct questions, and lower certainty on wrong questions, he received a positive bonus of 3.5% and his converted CB grade (4.4) is higher than it would have been with traditional accuracy (4.1).

To get some feeling for how the conversion from the CB grade to the converted CB grade works, we show some examples:

- C=high for all questions, 100% correct: 10 (you know you're correct)
- C=mid for all questions, 100% correct: 9
- C=low for all questions, 100% correct: 8 (underestimate your knowledge)
- C=low for all questions, 50% correct: 5 (you know you're guessing)
- C=mid for all questions, 50% correct: 4.5
- C=high for all questions, 50% correct: 3 (you don't know that you don't know)

With 50% of answers correct, whether the student passes or not depends on the confidence:

- 50% correct answers at C=high and 50% wrong at C=low: 6
- 40% correct answers at C=high and 10% correct at C=low and 50% wrong at C=low: 5.8
- 50% correct answers at C=mid, 50% wrong at C=low: 5.5
- 50% correct answers at C=low, 50% wrong at C=low: 5.0

Note however that the above items assume a correctly judged low certainty for all wrong answers (avoiding a negative bonus). As soon as you get questions wrong at mid or high certainty, it is possible to get a 5.0 even with 70% correct. (Note that this is a pathological worst case: 70% correct at C=low, 30% wrong at C=high. More realistic would be 60% correct, 40% incorrect, both at C=high, resulting in a 5.0.) Note that I would argue that it is fair to fail that with an accuracy higher than 50% or 55% when the student incorrectly judges the boundaries of his or her knowledge.

Research [5], illustrated in Figure 6, has shown that over time students judge their own knowledge more accurately, resulting in a shift from a spread of positive and negative bonuses to mainly positive bonuses in examinations. In other words, students receive a higher grade in CBM-based examinations than in traditional accuracy-only examinations.

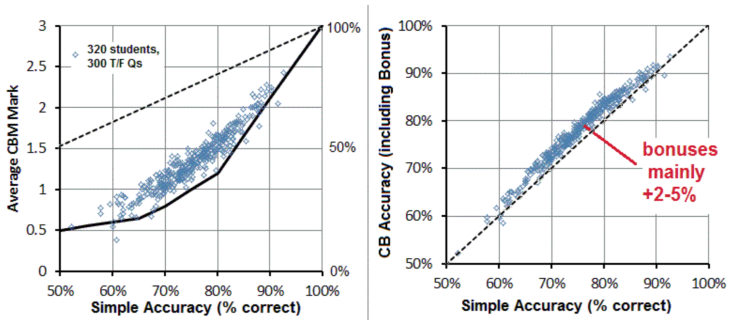

Figure 6: CBM bonuses in final examinations [5].

5 Experiments

5.1 Background

In the first quartile of the 2015/2016 academic year, we used CBL and CBM in 5EIA0 (Computation, 260 students) and 5AIA0 (Computation for Automotive, 110 students). In the first lecture we explained CBL and CBM. These two courses have almost identical course material, lectures, and practica. The intermediate exams were the same for both courses. The final exams different in 4 of 23 questions.

We enabled CBM in the Oncourse quizzes. For both homework and exams we used deferred feedback with CBM, which means that the student receives the grades and feedback (correct answer, explanation why the given answer is correct or not) after completing all questions in the homework or exam. For homework, an alternative is immediate feedback with CBM that gives feedback after every question. My (unsubstantiated) feeling was that this would make it too easy for students to not try hard enough and seek feedback (i.e. the right answer) immediately. In any case, this setting is trivial to change. The Oncourse log files show that students repeat their homework multiple times to improve their scores.

We used CBL and CBM:

- for the homework of weeks 2-6,
- for the five intermediate exams of weeks 2-6,
- for the practice examination in week 8,
- and also in the final exam in the examination week.

Students therefore had ample opportunity to get acquinted with this form of working and grading.

5.2 Data Analysis

Note that in the following analysis both accuracy and converted CB grade are on a scale from 0 to 10. The TUE final grade is from 1 to 10, obtained by multiplying with 9/10 and adding 1. This difference is not significant in the analysis that follows, which focusses on the relative difference between accuracy and converted CB grade.

Final Exams

The top histograms of Figures 7 and 8 show, for 5AIA1 and 5EIA1 respectively, the distribution of the CB bonus. The average bonus is -0.14 and -0.4, respectively. However, *60% of students had a positive bonus, i.e. they got a higher grade with CBM than with traditional grading.* The average negative bonus is thus higher (in magnitude) than the average positive bonus.

The lower histogram of Figures 7 and 8 show, for 5AIA1 and 5EIA1 respectively, the distributions of accuracy in blue and converted CB grade in red. The difference between the two is due to the bonuses.

The scatter plots of Figures 7 and 8 show, for 5AIA1 and 5EIA1 respectively, the converted CB grade (i.e. CB grade converted from -6..+3 to 0..10) on the Y axis against the accuracy (percentage correct questions) on the X axis. Points above the red line received a positive bonus, while those below the red line received a negative bonus. Starting from low accuracy (0-40%) we encounter students that don't know the answer and also mistakenly think that they do know the right answer. They receive a negative bonus, leading to the shift towards lower grades in the grade distribution shown above. Next, from 40-65%, we see students that increase their grade because they correctly indicate a higher certainty for correct answers and a low certainty for wrong answers (those above the red line). But also, students for which the converse holds (below the red line). However, we can observe a trend of increasing average bonus until around 80%. From 80% onwards, it is essentially only possible to get negative bonuses, which is what we observe. (Recall that 100% correct with C=low gives an 8, and 100% correct with C=high gives a 10.) Negative converted CB grades indictate students that received a very negative bonus (many wrong answers at high certainty); their grades are rounded up to 0.

Trends during the course

The bonuses in intermediate exams of weeks 2 to 6 are shown in Figure 9, 5AIA3 on the left and 5EAI3 on the right. The average bonuses for the intermediate exams for 5AIA3 were:+0.2 (60% positive) , -0.3 (40% positive), -0.6 (50% positive), -0.3 (44% positive), +0.3 (80% positive). The average bonuses for the intermediate exams for 5EAI3 were: +0.1 (66% positive), -0.2 (53% positive), -0.6 (40% positive), -0.2 (55% positive), +0.2 (73% positive). *Note that in 7 out of 10 intermediate exams 50% of students or more increased their score due to CBM.* Note that, for 5EIA3 at least, the obvious peak in certainty slowly moved from +0.1 in week 2 to +0.7 in week 6. The graphs for 5AIA3, traditionally weaker students than 5EIA3, do not show a clear trend.

The distributions of the accuracy (traditional grade) and converted CB grade of the intermediate exams of weeks 2 to 6 are shown in Figure 10, 5AIA3 on the left and 5EAI3 on the right. The first intermediate exam was much too easy and without any anti-fraud measures, resulting in very high grades (8.5 average) and positive bonuses. The third exam in week 4, contained errors and was too hard, as can be seen in the low (-0.6) average bonus. (We adjusted the grades for this week.) Apparently students were getting used to CBM and/or making more of an effort as the course proceeded, with bonuses increasing.

Figure 11 shows the converted CB grade on the Y axis against the accuracy on the X axis. There are few accurary levels (on the X axis) due to the limited number of questions in the intermediat exams, and that CBM provides a higher resolution, i.e. additionally distinguishes students by certainty (multiple points with same X value).

6 Notes on Oncourse and automated Grading

CBL works well when automated, such as when using an environment such as Oncourse (Moodle). In particular:

- Using quizzes for CBL homework, practice exams, and so on, allows students to take a quiz multiple times. Every time they take the quiz, they receive feedback on why their answers are correct or not, as well as whether they are under or over confident. In fact, the Oncourse log files show that students repeat their homework multiple times to improve their scores. For me, this is a major reason to use CBL.†
- Using quizzes for both homework, intermediate exams, and final exams gives students plenty of practice to work with Oncourse and CBM (if the latter is used).
- Although not intrinsic, by using CBL I paid more attention to wrong answers, and added explainations for common wrong answers.†
- Using quizzes to automatically grade homework and (intermediate) exams, clearly reduces the substantial grading workload of staff. *
- Using quizzes (Oncourse in general) for homework gives insight into what students do (or don't do), understand and find hard. We use this information to give feedback to students (e.g. do your homework), but also to see what topics have to be given more attention or repeated in class. †
- If quizzes are used for (intermediate) exams then the quiz can be opened for reviewing by students. As a result, students can review their own answers, get feedback on their (in)correct answers, and see their grade. This pretty much eliminates student questions on exams ("inzage"), and reduces the staff work at this state to comments about unclear questions. This saves a lot of time. *
- Automated grading eliminates questions about grading. However, note that using multiple choice questions with differential scoring per item (when some correct answers are worth more, or some incorrect answers are more wrong than others) gives rise to grading questions by students. The grade of each item should be given in the feedback per question to avoid this.
- Note that with automated grading, there is no leeway to increase grades by a notch to make students pass, as can be done when grading manually. This year I have students fail with 4.9 (when requiring a 5.0), where this was unlikely to have happened in previous years.
- Use of automated grading with Oncourse (and similarly PEACH) have raised some questions from students that automated grading of answers only is not fair and should always be backed up with a manual check of the

process/computation of how students have arrived at the answer. First, note that traditional paper multiple-choice exams are an accepted testing method (independently of whether they are manually or automatically graded), and that a post-hoc manual check is not required or even offered. Secondly, requiring a post-hoc manual check clearly entirely negates the point of automated testing.

† improve the quality of education we offer to students. * improve efficiency.

7 Student Evaluations of 5EIA0 (Computation) Course

The student evaluation of the course as a whole was broadly positive (Figure 12. Regarding CBM, 126 students answered the question "are you satisfied with certainty-based quizzes?" with an average of 2.8 on a scale from 1 to 5 (standard deviation=1.3; 40% with 1 or 2, and 28% with 3, and 32% with 4 or 5).

In free-form feedback, only one student explicitly mentioned that (s)he liked CBM. Nine students explicitly remarked that they did not like CBM, with the major complaint being that the conversion to the 1-10 scale from the CBM grade was unclear. Hopefully this can be easily rectified next year. As a side note, students were positive about the use of quizzes in general for homework.

To (cowardly) forestall any complaints about the use of CBM in the final exam, we used traditional accuracy as the grade for all students with a negative bonus, and the converted CB grade for all students with a positive bonus. This seems to have not been necessary, as only a few students of out 320 or asked about their grade.

8 Conclusions

Certainty-based grading (CBM) is a pedogagically interesting concept, aiming to quantify student insight into limits or certainty of their knowledge, and including it in the student's grade. Certainty-based learning (CBL) feeds this information back to the student such that he or she learns what topics he or she is knowledgeable about (i.e. correct with high certainty), or is uncertain about (i.e. correct or incorrect with low certainty), or is deluded about (i.e. incorrect with high certainty). CBM and CBL have been successfully used for many years at the University College of London (UCL).

CBM is intuitive and easy to explain, except for the conversion to traditional grading on a scale from 1 to 10. This hampers deployment, in the sense that it worries students. Some think that indicating the certainty requires some strategy, where in fact it doesn't (just be honest). Moreover, they agree with any grading scheme as long as they pass the course, and complain (about everything) as soon as they don't.

We introduced CBM and CBL to about 360 students in two courses. We then used CBL for homework (Oncourse quizzes with deferred feedback, with

unlimited number of attempts) and in intermediate and final exams (Oncourse quizzes with deferred feedback, with 1 attempt).

In the intermediate exams we observed that the average CB bonus increased (slowly) over time, as expected from prior research. In the final exams the average CB bonus was negative. However,*60% of students had a higher grade with CBM than with traditional accuracy grading.* In fact, *in 9 of our 12 exams at least 50% of students increased their grade with CBM.*

We observed that negative bonuses are concentrated on the lower half of accuracy (i.e. students with less than 65% accuracy), illustrating that these students were not aware that they were making mistakes.

Students have not complained about the use of CBM in the intermediate exams. To (cowardly) forestall any complaints about the use of CBM in the final exam, we used traditional accuracy as the grade for all students with a negative bonus, and the converted CB grade for all students with a positive bonus. This seems to have not been necessary, as only a few students of out 320 or asked about their grade.

Since there has not been major negative feedback from students we intend to continue with CBM and (especially) CBL at least for the homework and (intermediate) exams in other courses this year, and the same course next year.

References

[1] T. Gardner-Medwin, "Enhancing student learning and insight: using private self-tests with cbm," in *eLearning2.0*, 2015.

[2] Ahlgren, "Reliability, predictive validity, and personality bias of confidence-weighted scores," 1969.

[3] T. Gardner-Medwin, "http://www.tmedwin.net/ ucgbarg/pubteach.htm."

[4] ——, "Analysis of exams using certainty-based marking," in *Proc Physiol Soc series*, no. 3, 2006.

[5] ——, "Optimisation of certainty-based assessment scores," in *Proceedings of International Congress of Physiological Sciences*, 2013.

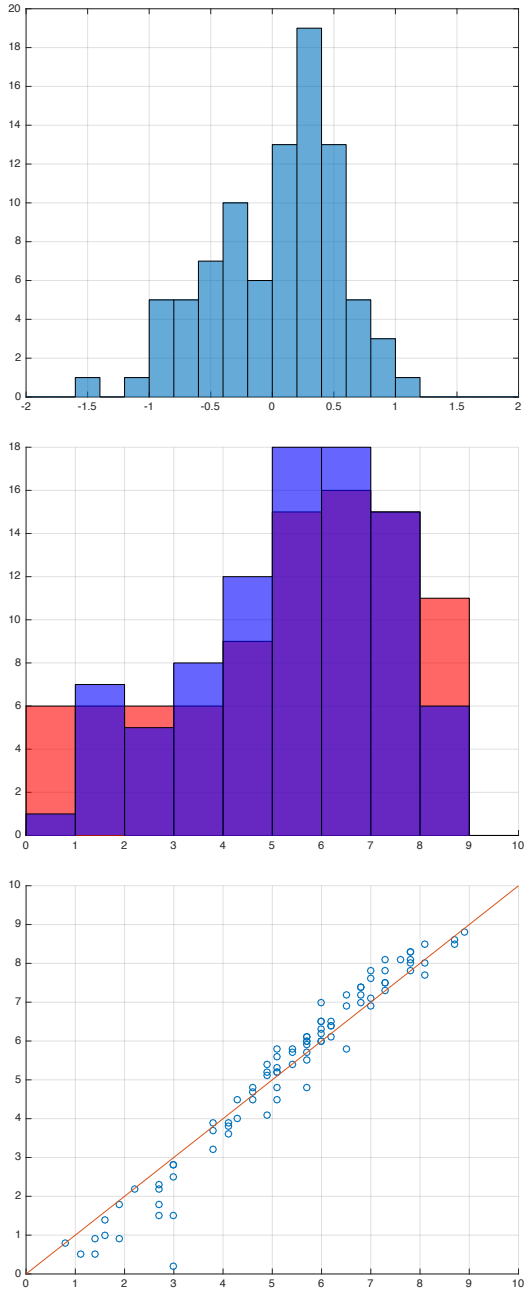

Figure 7: 5AIA1 final exam. From top: CB bonus, accuracy (blue) and converted CB grade (red), converted CB grade vs. accuracy.

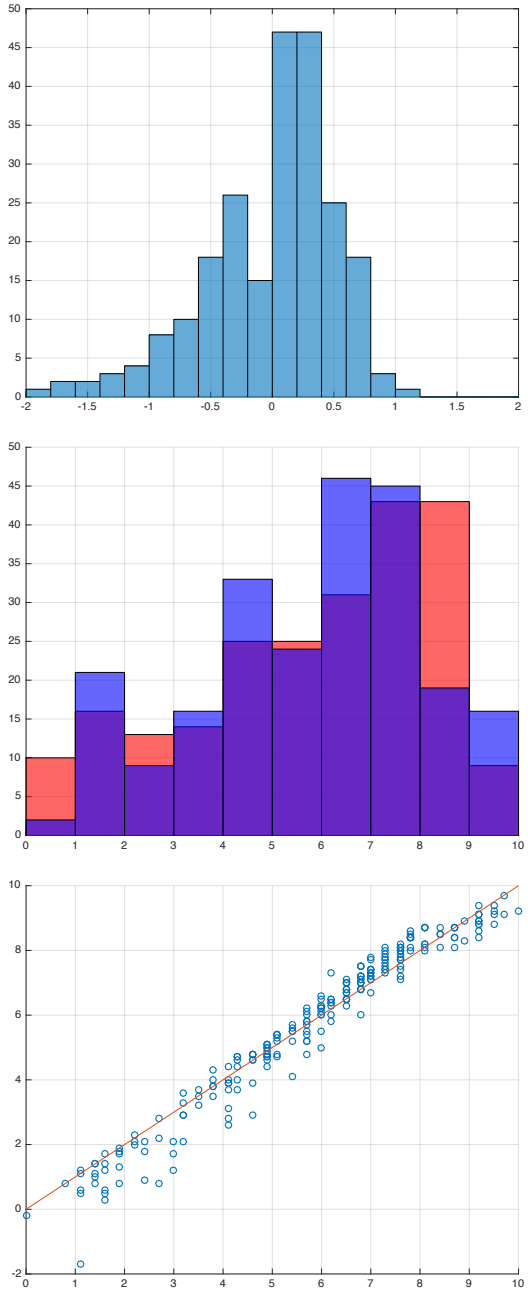

Figure 8: 5EIA1 final exam. From top: CB bonus, accuracy (blue) and converted CB grade (red), converted CB grade vs. accuracy.

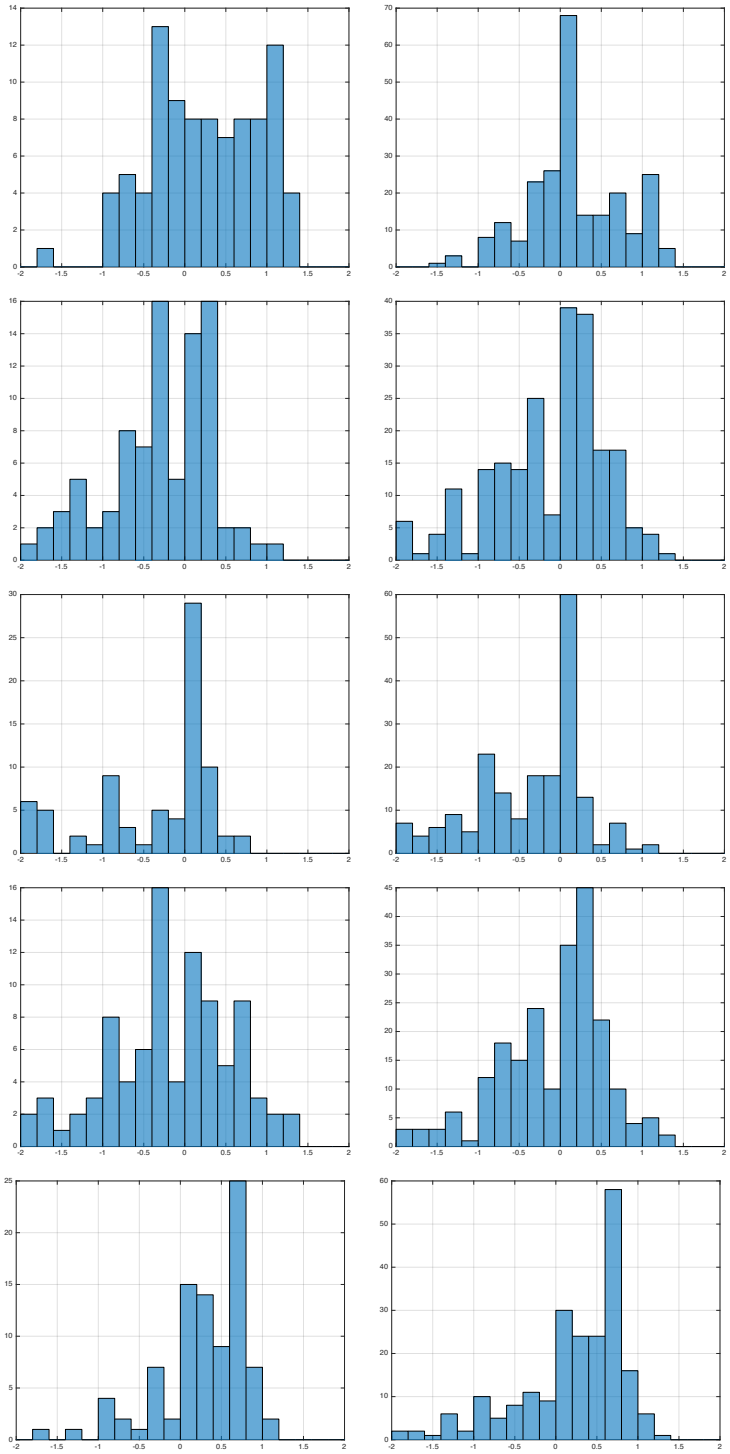

Figure 9: 5AIA3 (left) and 5EIA3 (right) intermediate exams. CB bonus.

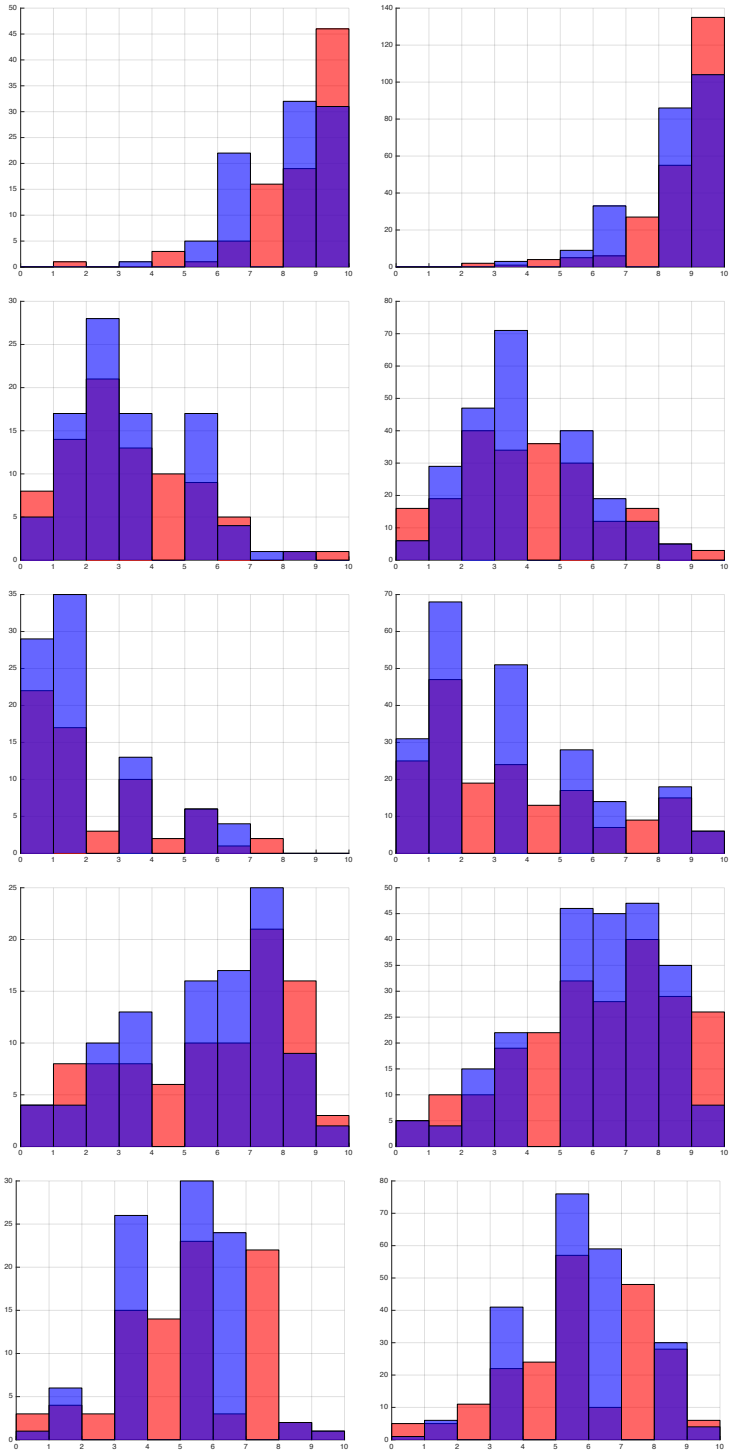

Figure 10: 5AIA3 (left) and 5EIA3 (right) intermediate exams. Accuracy (blue) and converted CB grade (red).

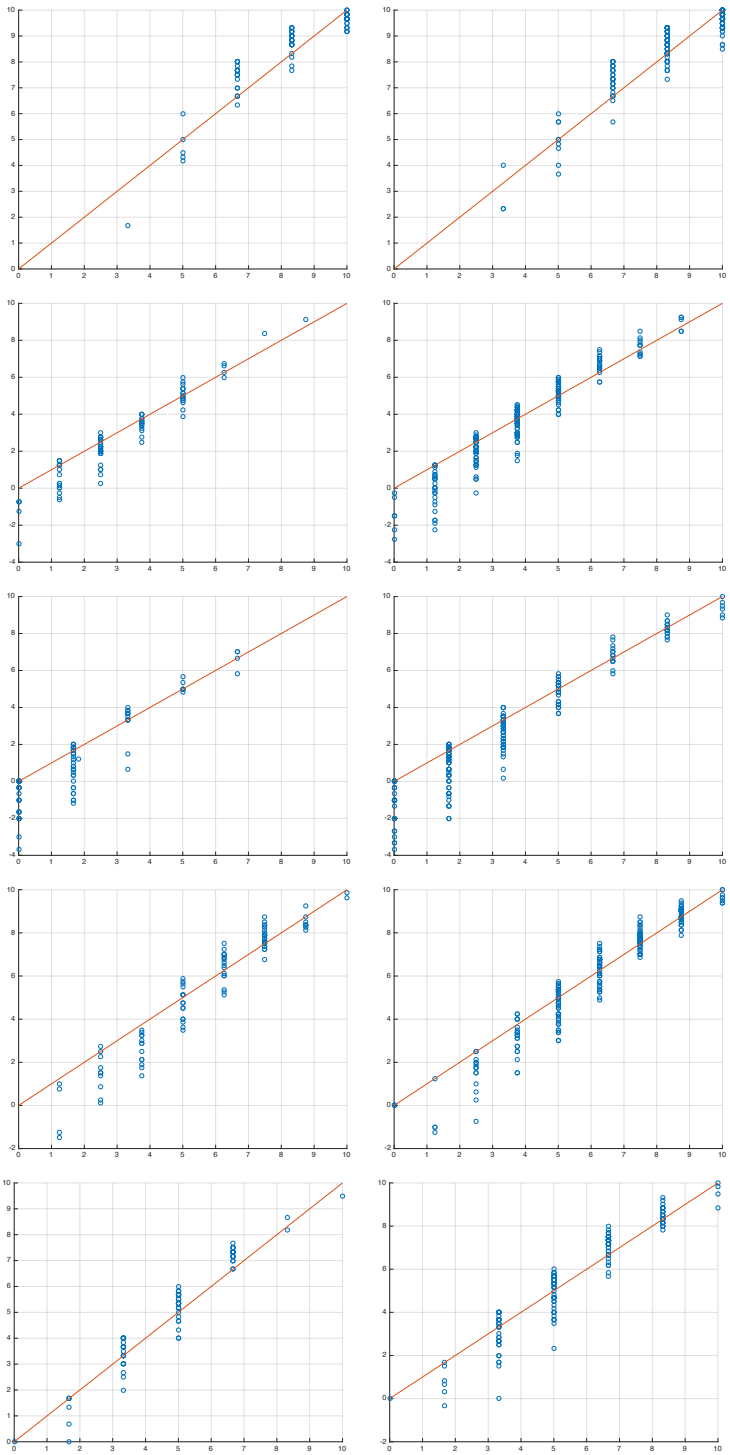

Figure 11: 5AIA3 (left) and 5EIA3 (right) intermediate exams. Converted CB grade vs. accuracy.

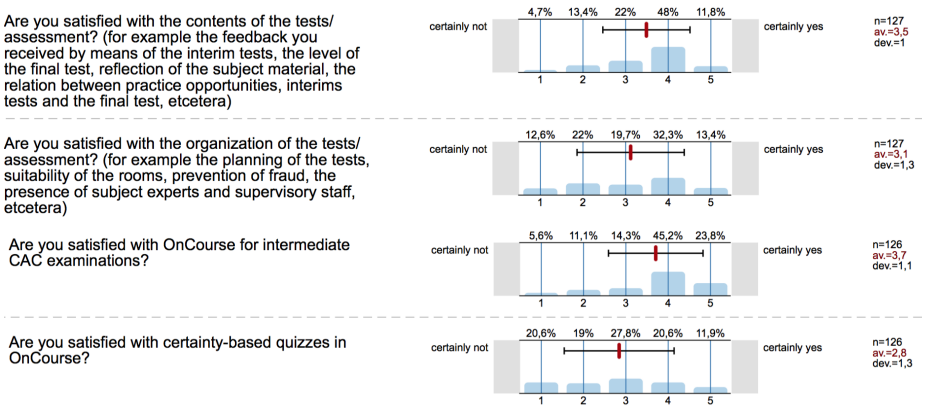

Figure 12: Student feedback on the 5EIA0 course.