How Multiple Choice Exams are Graded

1. Rationale:

Let's consider a multiple choice exam where, for each question, ***K*** answer options are presented, and the following set of conditions are assumed :

- one and only one of the answers is correct;
- there are no other valid answer options besides the one that is correct;
- the distribution of correct answers among the ***K*** options, for the entire exam, is uniform.

Under these conditions, it is possible to respond to the entire exam ignoring, a priori, the entire nature of the knowledge under evaluation. In the limit, it is even possible to answer the entire exam without even reading the presented questions.

Therefore, for someone being evaluated, there is always the possibility of playing with probabilities to obtain a rating above zero by randomly filling in the response matrix.

The probability of getting the right answer, playing exclusively with luck, is given by

$P_c = \frac{1}{\boldsymbol{K}}$ while the probability of failing is given by $P_w = \left(1 - \frac{1}{\boldsymbol{K}}\right) = \frac{\boldsymbol{K}-1}{\boldsymbol{K}}$

Thus, if the rating assigned to a correct answer is $\boldsymbol{V_c}$, and the rating assigned to a wrong answer is $\boldsymbol{V_w}$, then, in order for the result to be statistically null after playing only with luck, we can conclude that:

$$V_c * P_c + V_w * P_w = 0 \qquad \Rightarrow \qquad V_w = -\frac{V_c * P_c}{P_w} = -\frac{V_c}{(k-1)}$$

As an example, and in the case of a Computer Architecture I theoretical exam, there are 4 answer options per question (***K*** = 4). Therefore, the grading to be assigned to the wrong answers should be

$$V_w = -\frac{V_c}{3}$$

2. Assumption

The rationale expressed in point 1 assumes that, whoever takes the exam, has no knowledge on the topics in which he is being evaluated. This may, however, not be the case for a significant number of those students undergoing evaluation. In that case, penalizing wrong answers based on the assumption that they result from a game-based answer may no longer be reasonable. This raises the question of which method and criteria to adopt in cases where it is reasonable to consider that the answers given are not the result of chance alone.

A good starting point will be to assume that, regardless of the number of questions actually answered, the ratio between the number of wrong answers and the number of right answers will grow smaller with the increasing knowledge and confidence placed in the answer selections. In this assumption, we can consider the following theoretical limits:

Let $\boldsymbol{N_w}$ be the number of wrong answers and $\boldsymbol{N_c}$ the number of right answers. The ratio

$$F_c = \frac{N_w}{N_c}, \text{ limited on the left to } [K-1]$$

represents de confidence level on the actual knowledge of the evaluated subject, assuming that, for K answer options,

$$F_c \in [K - 1 \dots 0]$$

and also assuming that the probabilities expressed at point 1, for an exclusively random answer is statistically applied on average.

The confidence factor will therefore be higher for a lower value of F_c , regardless of the number of questions the subject answers. An evaluated student who, for example, answers only half of the questions, but get them all right, will reveal a maximum confidence factor (0) for the questions he chose to answer.

Under this assumption, a high confidence factor will be assumed to be associated with a good knowledge of the topics answered by the person being evaluated, indicating that the wrong answer(s) should be due less to a bet on luck and more to a factor of lack of knowledge, or due to an error in the evaluation of the question.

On the other hand, when the confidence factor decreases, it is legitimate to assume that part, or all of the given answers, were due to a bet on the luck factor and not on effective knowledge on the evaluated topic.

3. Negative grading corrected according to the confidence factor

Taking into consideration the conclusions of the previous point, the confidence factor starts to be used as an element of correction to the quotation attributed by a wrong answer ($\boldsymbol{V_w}$).

The correction is obtained by multiplying the value of $\boldsymbol{V_w}$ by a correction factor $\boldsymbol{C_F}$ given by:

$$C_f = \left(1 - e^{-((F_C - 0.15)\,*\,4)}\right)$$

$$\text{with } C_F = \begin{cases} 0 \; if \; C_f \leq 0 \\ \quad C_f \end{cases}$$

In the next figure the C_F function is displayed for values of $F_c \in [0 \ldots 3]$

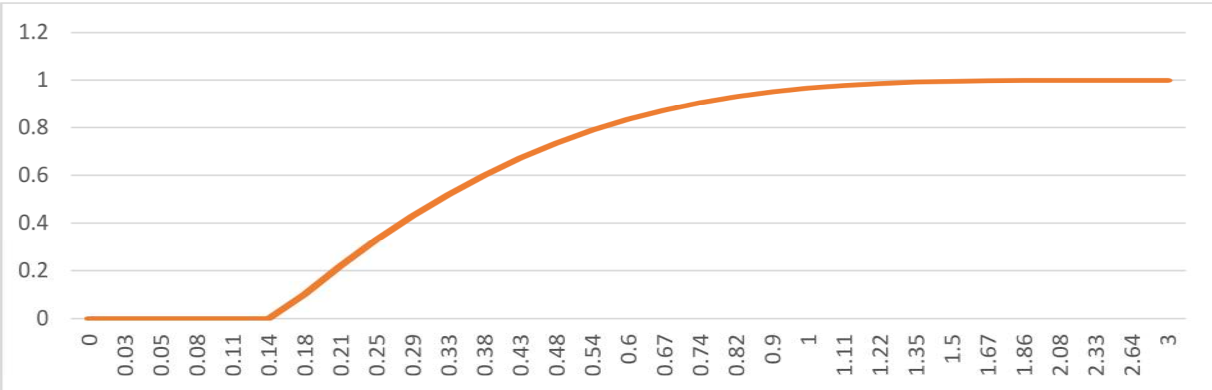

Fig. 1 – C_F function for values of F_c ϵ[0 .. 3]

Example:

Exam with 40 questions and 4 answer per question:

Number of total answers provided by the evaluated subject: 32

Number of right answers: 26

Number of wrong answers: 6

Quotation of each question: 0.5 out of 20.0

- Grading resulting from the exam correction without the confidence factor:

$$Final\ grade = (26 * 0.5) - 6 * \frac{0.5}{3} = 12.0\ out\ of\ 20.0$$

- Grading after inclusion of the confidence factor:

$$V_w = -\frac{1}{3} \qquad F_c = \frac{6}{26} = 0.2308 \qquad C_f = \left(1 - e^{-((0.2308 - 0.15)\,*\,4)}\right) = 0.2761$$

$$Final\ grade = (26 * 0.5) - 6 * \frac{0.5}{3} * 0.2761 = 12.72\ out\ of\ 20.0$$

4. Negative grading in cases where double answers are used (selection of two answers out of *K*)

Let's assume that:

- $D_f = \frac{C_f}{k-1}$ is a discount factor dependent on the confidence level;
- N_{od} is the number of questions to which the subject has answered with a a double option;
- M_{od} is the maximum number od double answers allowed in a particular exam;
- V_c the grade applied per each of the correctly answered questions;
- $T_D = \frac{V_c}{k-1}$ the maximum negative grade to be applied.

For those cases where one or more double answers are used, the classification is obtained by using the D_{BD} value, obtained as follows:

$$D_{BD} = ((0.1 + D_f) * \left(1 + \left(\frac{N_{od} * 4}{M_{od} * (k-1)}\right)\right)) * V_c$$

$$\text{being } D_{BD} = \begin{cases} T_D \; se \; D_{bd} > T_D \\ D_{bd} \end{cases}$$

The rationale behind this approach stems from the observation that resorting to a double answer, even if one of the options is correct, is not the same as correctly answering the question using only one option. Thus, although the confidence factor is also applied in this case, there will be a minimum amount of discount to be applied to the wrong option.

More specifically, this minimum discount corresponds to the case where only a single double answer is used and one of the selected options is correct.

Example:

- Maximum of 6 double answers possible;
- Only one question has been answered with a double option;
- For the maximum confidence level (= 0);

$$D_{BD} = \left((0.1 + 0) * \left(1 + \left(\frac{1 * 4}{6 * (4-1)}\right)\right)\right) * V_c = 0.1222 * V_c$$

If V_c = 0.8 then

$Grade = 0.8 - (0.8 * 0.1222) = 0.7022$ if one of the options is correct, or
$Grade = -2 * (0.8 * 0.1222) = -0.1955$ if both options are wrong

- For the minimum confidence level (= 3);

$$D_{BD} = \left((0.1 + 3) * \left(1 + \left(\frac{1*4}{6*(4-1)}\right)\right)\right) * V_c = 3.788 * V_c \;\; >> \;\; D_{BD} = \frac{1}{3}$$

If V_c = 0.8 then

$Grade = 0.8 - (0.8 * 0.3333) = 0.533$ if one of the options is correct, or
$Grade = -2 * (0.8 * 0.3333) = -0.533$ if both options are wrong