

GRADE LEVELS FOR MULTIPLE-CHOICE TESTS

Per Westerlund¹, Patrik Hilber¹ and Tommie Lindquist²

¹ Department of Electromagnetic Engineering,
School of Electrical Engineering,
KTH Royal Institute of Technology, Sweden

² Svenska kraftnät (Swedish National Grid), Sweden

In education there is nearly always an evaluation of how much of the course syllabus the student has learnt. The multiple-choice tests are a type of written examination that is easy to use, since the correction is quickly done. The problem with these test is to find a good grading scale, since the students can get more points than what their knowledge indicate, as they can guess.

There are different approaches to analyze the results of multiple-choice tests. However, the articles that have been found do not discuss the number of grades that a multiple-choice test should be able to differentiate. Also in those articles there are no figures, where the effect of guessing is shown.

The purpose is to design a multiple-choice test. The inputs are the grade scale and the number of alternatives and the output is the number of questions required for a given level of errors in grading. As an example we consider a multiple-choice test with 40 questions with 4 alternatives. We study the distribution of four students that knows nothing, 25 %, 50 % and 75 % of the questions. They guess the rest of the questions. The points of each student will follow a binomial distribution as in figure 1.

It can be observed that the distribution narrows when going from the left to the right. This is due to a smaller variance, since the number of guesses decreases. With a smaller variance, the height of the peak has to increase, in order to get the area below the curve. The overlapping tails are bigger to the left, as the variance is bigger. Thus the risk of putting a student in the wrong category is bigger for the lower grades. The rightmost peak is the least symmetric, with a steeper left-hand side. Then the binomial distribution is not approximated by a normal distribution so well, since the number of trials is lower. Figure 2 is the same as figure 1 except that there are 60 questions. The intersection between the curves is at a lower level of probability, which means that the risk of giving a student a wrong grade is lower.

Keywords: Binomial distribution, normal approximation, multiple-choice tests.

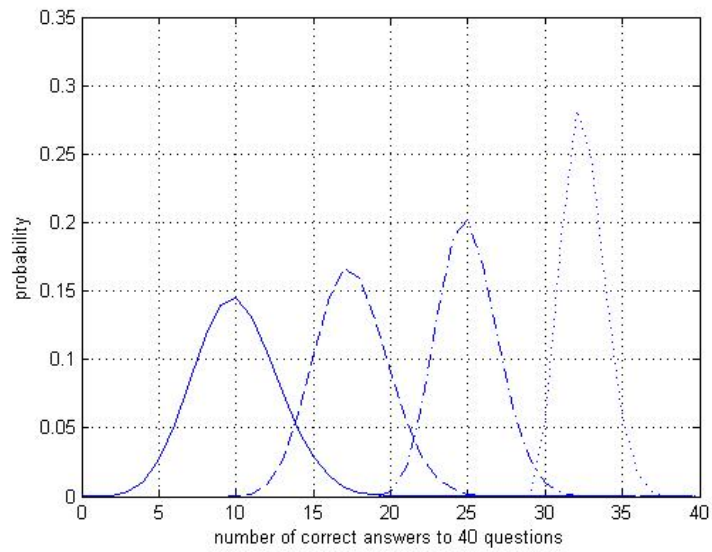


Figure 1: Probability distribution for 40 questions with different levels of knowledge (solid: 0, dashed: 25 %, dashdot: 50 %, dotted: 75 %).

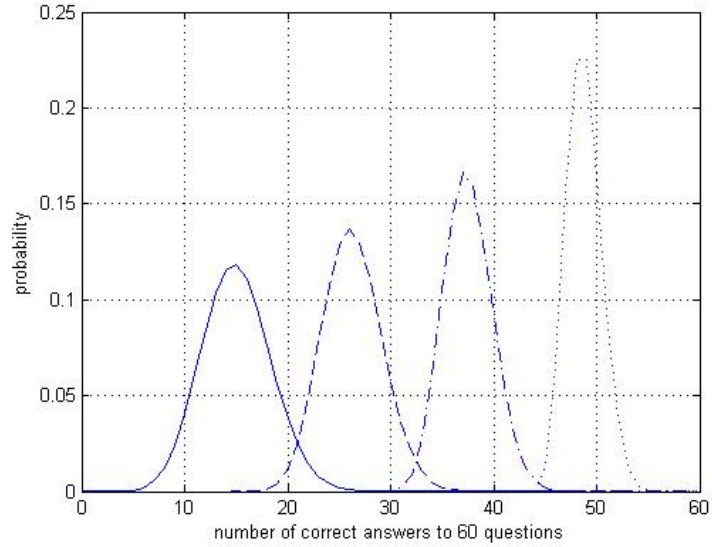


Figure 2: Probability distribution for 60 questions with different levels of knowledge (solid: 0, dashed: 25 %, dashdot: 50 %, dotted: 75 %).