

QUESTION RANKING METHODS BY DETERMINING THEIR DIFFICULTY

Alin ZAMFIROIU

Bucharest University of Economic Studies, Bucharest, Romania
National Institute for Research and Development in Informatics, Bucharest, Romania
alin.zamfiroiu@csie.ase.ro

Ramesh C. SHARMA

Instructional Design at Dr B R Ambedkar University Delhi, New Delhi, India
rc_sharma@yahoo.com

Abstract. *Assessment of learning is an important aspect of educational pedagogy. Right kind of assessment strategies help the teachers obtain right kind of feedback. One of the strategies for assessment is to deploy quizzes for formative as well as summative assessment. Suitably designed quizzes prove to be an effective tool for this purpose. With this objective, we propose to develop an algorithm to determine the quality of the quiz questions.*

Many assessments are made through quiz questions. But their difficulty level is determined by the person who carries out these questions, the teacher. The algorithm that we propose in this material establishes the difficulty level of the quiz questions according to the answers of those to whom these questions are addressed, i.e. the students.

In the past, many platforms have an approach to supplying the questions in increasing difficulty. In this way it offers the student the possibility of gradual evolution. The difficulty of the questions is calculated based on the answers of other students in the preparation phase of the questions.

Out of several sets of questions we did with the students, we got a set of 22 questions with 47 answers. Based on these answers we calculated the difficulty of these 22 questions. The algorithm can be applied to others sets of quiz questions.

Keywords: assessment, algorithm, e-learning, quiz, questions, difficulty

JEL classification: I21 Higher Education • Analysis of Education

DOI: 10.24818/ie2020.05.01

1. Introduction

According to [1] the instructors and teachers are confronted with a big challenge during the student learning process. They have to determine the conditions and material of instruction that optimize the student learning and also their evaluation. Evaluation should be equitable for all students and should be progressive.

The aim of this paper is to establish the quality in terms of difficulty of quiz questions by assessing the students' knowledge and their interaction with these questions in the learning management systems.

Currently, there are numerous applications in the educational field that help to improve the academic performance of students, applications that use digital rankings or badges to help engage, motivate and maintain student attention. Over time, there has been a positive relationship between student performances (using these educational IT systems) and obtaining higher grades. Most test applications (whether mobile, web, or desktop) are for higher education and have been designed to address a key aspect of the learning process, such as

collaboration or motivation. At this time, student demand for personalized education is increasing due to competitiveness because the tests generate feedback, allowing students to see how they behave compared to their peers, stimulating the need for self-realization.

There are various online solutions for the evaluation of students. Some of them are analysed in next paragraphs.

Socrative [2] is a free student evaluation system available from any device connected to the Internet (tablet, smartphone, laptop) through web applications (Chrome, Safari, firefox etc) or mobile (for iOS or Android), which allows teachers to create and distribute tests to the students with different types of questions over a designated time duration: true / false, multiple choice, short answer questions.

QuizCreator [3] is an online test software system that allows users to build tests and surveys, and then track results. This system is composed of two applications, namely: a QuizCreator desktop application in which the actual tests are built and a QuizCreator Online site in which the test is published and the results can be visible.

Easy Test Maker [4] is an easy-to-use student assessment tool, presented in a simple and highly organized format. This application allows the creation and publication of tests online.

Kahoot [5] and Quizizz [6] are other applications used for quizzes. The advantages of these applications are that the teacher can establish the time for each question. In this way the students will have limited time for each question to answer. This factor may reduce the chance for cheating.

For all these applications it is very important to add a new functionality and to consider the difficulty of the questions from the quiz. In order to calculate the difficulty of quiz questions it is pertinent to use these questions for other students and see the reaction of them regarding to the questions.

In [7] the IRT – Item Response Theory is presented. Our method is based on the same theory and it is necessary that the student to answer the quiz questions in advance. Our model proposes that once a different student answers it in the ranking process, and after that the questions it will be used for other students. Also, in [7] there is another model of estimating the difficulty level without answering to the quiz questions. They analyse the quiz questions types to estimate the difficulty level.

In [8] the authors suggest that an ideal method for ranking the questions is to take an arbitrary three questions and establish an ordering rule between them. For validating they have used 60 sets of questions. Another research highlights an approach of ranking the questions by its difficulty [9]. In this paper the authors are using a differential equation and developed an Intelligent Tutoring System for mathematics education: MathITS. In this ITS, the students are able to complete tests and after that they get a constructive feedback with their weaknesses indicated and the concepts that should be prepared more, until next test.

2. Method

2.1 Questions

The bank of grid questions to be evaluated is created. These can be created by the teacher, or they can be obtained by students who have passed the content of the subject. All the obtained questions are saved and a question bank is created on the MOODLE platform. On the platform we can create more questions banks and in that way we can use all of them anytime that we need it.



Figure 1. Questions bank

2.2 Online quizzes

With this question bank, a quiz is created on the MOODLE platform where students can enter and take the grid test whenever they want.

The respondents who take part in these grid tests are those who have experience in the field from which the respective questions are created and have a constructive attitude towards the question bank.

2.3 Measures

All answers to the questions are saved for future analysis. Both the answers from the users' first support for each question and the answers from the subsequent support are saved. Thus, an analysis can be made regarding the evolution of the responses of some participants over time. If some questions have a very high degree of difficulty, the participants will not answer correctly from the second / third attempt either.

In addition to the answers of the questions, the final results of the users will be saved. This will allow to calculate the difficulty of the questions based on their knowledge. If a user has answered to more questions correctly and has only a few incorrect answers, these questions have a high level of difficulty.

Also, the time spent by the user during the test is saved, so that the degree of confidence of the users in their answers is also used [10].

2.4. Procedure

When the results of the tests are obtained, they can be analyzed and several methods can be used to determine the difficulty of the questions.

Within this framework, we propose two methods. The first method involves counting all the answers to a question and making a report with the total number of respondents who answered that question. The used formula is:

$$QD_i^1 = 1 - \frac{\sum_{j=0}^m A_{ij}}{m}, i = \overline{1, n} \quad (1)$$

where:

- A_{ij} – the answer given by participant j for question i . The possible value is 0 if the answer is incorrect or 1 if the answer is correct;
- m – number of students participating in the test;
- QD_i^1 – the difficulty of the question calculated using method 1.

The second method also takes into account the final result obtained by the respective participants. In this way, the difficulty of the question is also influenced by the general knowledge of the test participant.

For this method the applied formula is:

$$QD_i^2 = 1 - \frac{\sum_{j=0}^m A_{ij} * (R_j / NQ)}{m}, i = \overline{1, n}$$

where:

- R_j – represents the result obtained by the participant j in the test;
- NQ – represents the number of questions in the question bank.

By applying these two methods two sets will be obtained with the difficulties of the questions. These sets can be taken into account for scoring questions in future tests for participants.

3. Results

For the validation of these two methods, a bank with 22 questions was prepared. These questions were collected from the students and it was desired to carry out a test with some of them. But for this the difficulty of each question must be determined and then those with high difficulty will be selected for inclusion in the final test.

A grid test was performed, which has been accessed 47 times. Thus, there were 47 supporters of the test. Among these 47, some were made by the same students. For our analysis we considered all 47 records. Applying the first method of calculating the difficulty of the questions, the difficulty levels from *Table 1* were obtained.

Table 1. Difficulty grades for questions determined with the first method

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
0.04	0.13	0.33	0.17	0.04	0.20	0.00	0.20	0.54	0.35	0.15
1.00	0.24	0.30	0.13	0.07	0.15	0.02	0.61	0.13	0.07	0.33
Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22

The limit cases for this analysis are represented by question Q7 which obtained the difficulty level 0 and the question Q12 which obtained the difficulty degree 1. Question Q7 obtained 0, because all the answers to this question were correct and for Q12 it was not obtained. no correct answer, and for this reason the degree of difficulty is 1 - the highest.

By applying the second method, the difficulty levels obtained have been show in *Table 2*. Within this method, the final score obtained by each participant was taken into account.

Table 2. Difficulty grades for questions determined by using the second method

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
0.26	0.31	0.46	0.36	0.26	0.37	0.24	0.35	0.63	0.48	0.34
1.00	0.41	0.46	0.31	0.29	0.33	0.24	0.67	0.32	0.26	0.45
Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22

It can be observed that in this method the question Q7 (which through the first method had difficulty level 0), now has 0.24, is still the lowest difficulty, but the students' results influenced the level of difficulty obtained.

It can also be observed that the trend remains for the majority of the questions regardless of the method applied. This is also shown visually in *Figure 2*.

As can be seen from the figure, the obtained values by the second method are higher, with an average of 0.4 compared to 0.24, the average of the values obtained by the first method. The advantage of the second method is that the values obtained by this method have a standard deviation of 0.17, while the values obtained by the first method have a standard deviation of 0.23. Thus, the values for the difficulty of the questions obtained through the second method are closer.

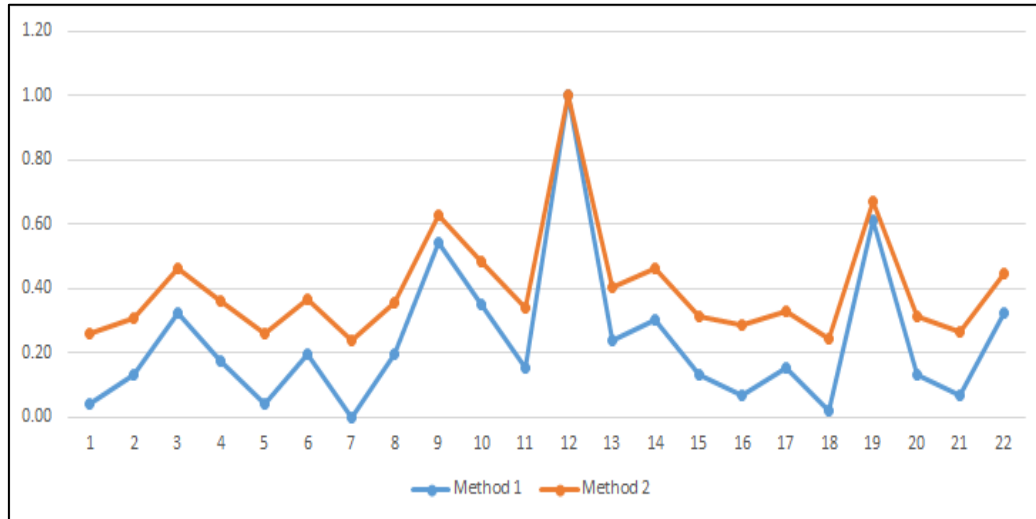


Figure 2. The degrees of difficulty obtained by the two methods

5. Conclusions

These two methods of determining the difficulty of the questions within the tests can be used for any question bank. To these two methods of calculating the difficulty can be added and other methods that take into account a single student attempt, or the time spent in the test by each participant.

We, further, want to extend this algorithmic process in order to add other methods to calculate the difficulty of the questions and to validate these methods on several question banks.

Acknowledgments

This paper was co-financed from the Human Capital Operational Program 2014-2020, project number POCU / 380/6/13/125245 no. 36482 / 23.05.2019 "Excellence in interdisciplinary PhD and post-PhD research, career alternatives through entrepreneurial initiative (EXCIA)", coordinator The Bucharest University of Economic Studies".

References

- [1] E.L. Bjork, J.L. Little and B.C. Storm. "Multiple-choice testing as a desirable difficulty in the classroom", Journal of Applied Research in Memory and Cognition, vol. 3, no.3, pp. 165-170, 2014.
- [2] Socrative, Online: <https://socrative.com/>
- [3] QuizCreator, Online: <http://www.quiz-creator.com/online-quiz-maker/>
- [4] Easy Test Maker, Online: <https://www.easytestmaker.com/>
- [5] Kahoot, Online: <https://kahoot.com/>
- [6] Quizizz, online: <https://quizizz.com/admin>
- [7] S. Ikeda, T. Takagi, M. Takagi and Y. Teshigawara, "A Study on a Method of Estimating the Difficulty of Quizzes Focused on Quiz Types", Proceeding from ICCE2011, The 19th International Conference on Computers in Education, Chiang Mai, Thailand, pp. 312-316, 2011.
- [8] R. Bunescu and Y. Huang, "A utility-driven approach to question ranking in social QA", Proceedings of the 23rd International Conference on Computational Linguistics, pp. 125-133, Association for Computational Linguistics, 2010.

- [9] K. Gunel and R. Asliyan, "Determining difficulty of questions in intelligent tutoring systems", Turkish Online Journal of Educational Technology-TOJET, vol. 8, no.3, pp. 14-21, 2009.
- [10] A. Zamfiroiu, A., I. Petre and R. Boncea, "Quiz Logs Analyse on Moodle E-Learning Platform", International Open and Distance Learning Conference Proceedings Book, p. 51, 2019