Open Peer Review on Qeios

Improving Learning Outcomes through Well Designed MCQ Tests

Nishtha Phutela[1], Atul Mishra[1], Kiran Sharma[1], Yogesh Gupta[1], Kushal Shah

1 BML Munjal University

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

The rapid growth in the number of students pursuing engineering degrees worldwide has resulted in an overburdening of teachers at many universities and engineering colleges. This situation makes it challenging for them to maintain high standards of assessments and provide personalized feedback to students. In large classes, many teachers use MCQ-based examinations for assessments, but these exams suffer from high incidences of copying and guesswork by students. In addition, large class sizes also make it difficult for the teacher to provide effective personalised feedback to each student. To address both these issues, we have developed a specialized protocol for conducting online MCQ-based examinations. We tested this protocol in a few courses at two universities, and our findings suggest that its use can significantly reduce incidences of copying and guesswork in MCQ-based assessments, and also help in providing personalized feedback to each student for further improvement.

Nishtha Phutela[1,a], **Atul Mishra**[1,a], **Kiran Sharma**[1,a], **Yogesh Gupta**[1,b], and **Kushal Shah**[2,c]

[1] *Computer Science and Engineering Department, BML Munjal University, Gurugram, India*

[2] *Sitare University, India*

[a] Assistant Professor at BML Munjal University.

[b] Associate Professor at BML Munjal University.

[c] Faculty member at Sitare University, visiting faculty at BML Munjal University and also the founder of Self Shiksha and Pariksha.

Keywords: Education, assessment, feedback.

I. Introduction

Jacques Barzun once rightly said, “Teaching is not a lost art, but respect for teaching is a lost tradition.”[1] Now while it is true that teaching is not a lost art, teachers have surely been facing difficulties in adapting well to address current

challenges [2][3][4]. Due to a growing population and higher monetary benefits of pursuing an engineering degree, the number of students enrolled in various engineering colleges have become very large [5]. Now while the number of engineering students has grown, the number of teachers who can teach these courses have not and perhaps cannot grow proportionately. And the end result has been that most engineering classes are very large in size, thereby making it very challenging for teachers to maintain the same quality of teaching.

Large class sizes mainly pose two challenges. For teaching, it becomes difficult to keep track of individual student's progress and provide meaningful feedback [6][7]. For assessments, it becomes difficult to have high quality subjective exams, since it will take too much time to manually correct the answer scripts. As a result, a lot of students feel disengaged from the courses, leading to poor learning outcomes. A certain percentage of highly motivated students are still able to make up for this loss through online resources, but a significant fraction is just expected to be happy to have obtained an engineering degree without knowing much about its concepts [8].

When the pandemic struck in 2020, several EdTech companies saw a huge surge in business, and it felt as if the age of online education has finally begun. But fortunately or unfortunately, what we clearly saw during the pandemic was that our current online education systems are not very efficient for the vast majority of undergraduate students. This is why most universities have now rightly gone back to offline teaching. So what is badly needed is a technological solution that can help our teachers in effectively teaching large classes and ensuring good learning outcomes for all students, without any exception.

In this paper, we present a new methodology of conducting MCQ [Multiple Choice Questions] tests which can help in significantly improving the effectiveness of teaching and assessments. Now, teaching has three components. First is lecture delivery, second is assessment and the third is personalised feedback. Lecture delivery is not really a problem since all that is required is a large enough classroom and big enough projector screen, which most universities have. However, providing meaningful personalised feedback becomes a challenge in large classes. When it comes to assessments, a major challenge with large classes is copying. And if the exam is MCQ based, another major problem is guesswork. Due to these two problems, the marks distribution of MCQ based exams is generally found to be very skewed towards the higher side (i.e. has a high mean and low standard deviation).

Past research has shown that choosing the MCQs randomly from a large enough question bank can adequately address the problem [9]. In particular, as shown in Table I, it was found that increasing the randomization in the questions to around 80% helps in achieving the desired marks distribution in a class [9]. Here, 80% randomization means that 80% of the questions seen by any student would be unique and not seen by any other student. However, preparing such a large question bank for a faculty member is very difficult for most courses.

Table I. Effect of high randomization in MCQ tests [9]. Here, 80% randomization means that 80% of the questions seen by any student would be unique and not seen by any other student. It can be clearly seen that a higher amount of randomization can help in achieving a good marks distribution even for MCQ exams.

Exam Type	Mean Marks	Standard Deviation
No randomization	96.36	3.4
80% randomization	61.80	12.5
92.5% randomization	60.0	9.9

Through controlled experiments in the academic courses being offered at two universities (BML Munjal University and Sitare University), we have demonstrated that an intelligently designed MCQ test methodology can solve both these objectives of providing personalised feedback and reducing copying as well as guesswork. In the next section, we discuss about Outcome Based Education (OBE) which is the concept utilizing quantitative measurement of student learning outcomes and plays a pivotal role in conceptualization of our proposed MCQ exam protocol.

II. Outcome Based Education

Outcome Based education (OBE) is a systematic way of designing the curriculum and assessments around specific learning outcomes, which helps in a quantitative analysis of student performance in a detailed manner and in providing personalized feedback to each student for further improvement [10]. Essentially, it begins with the teacher or curriculum designer identifying specific and measurable learning outcomes that are to be achieved after the completion of the course, and then quantitatively measuring the extent to which each learning outcome has been achieved.

Education methods based on the OBE principles have been adopted at multiple levels in various education systems around the world. In some countries, OBE based methods have been successfully working since the last few decades, whereas in a few other countries, they were abandoned after initial implementation [11]. As per our understanding, the resistance from teachers in a few countries was mainly due to poor implementation. To facilitate a wider international adoption of OBE based methods, the Washington Accord was created in 1989. It is an agreement by the signatory countries to accept undergraduate engineering degrees that were obtained using OBE methods. Several countries are its signatories, including the United States, the United Kingdom, Russia, South Africa, China and India [12].

In the Indian context, Outcome Based Education [OBE] is also known as Competency Based Education [CBE], and is an important part of its National Education Policy [NEP] 2020. This can be a game changer for all students if implemented effectively. In particular, the NEP 2020 document [13] released by the Government of India says, “The aim of assessment in the culture of our schooling system will shift from one that is summative and primarily tests rote memorization skills to

one that is more regular and formative, is more competency-based, promotes learning and development for our students, and tests higher-order skills, such as analysis, critical thinking, and conceptual clarity. The primary purpose of assessment will indeed be for learning; it will help the teacher and student, and the entire schooling system, continuously revise teaching-learning processes to optimize learning and development for all students".

There are, of course, a lot of ways in which OBE can be implemented, but a simple and practical implementation involves three steps:

1. Define Course Learning Outcomes [CLOs].
2. For all exam questions, identify the CLOs and the Bloom's Taxonomy Level [BL][14].
3. Evaluate the performance of the class and each individual student using the CLO and BL information, and enter the marks obtained by each student for each question in a spreadsheet.

This spreadsheet can then be uploaded on a suitable software platform for further analysis. A detailed data collection in this format helps in carefully analyzing the kind of mistakes students make, and also in getting a detailed understanding of the strengths and weaknesses of students across the country. For example, we will get to know what fraction of students are overall weak in most of the CLOs and BLs, how many are weak only in BL1, how many are weak only in CLO3, and so on. These statistical insights are very helpful in examining the overall class performance in detail, as well as providing personalized feedback to each student for further improvement.

III. MCQ Exam Protocol

Although the OBE approach described in the previous section can be used for subjective examinations as well, our focus in this paper is to use it for MCQ based exams. Now, for this process to be effective, we need to ensure that copying in these exams is reduced since otherwise the statistical results obtained from these exams will not be reliable.

Through detailed discussions with various faculty members at BML Munjal University, we have arrived at the following list of features that a robust online MCQ exam software platform should have:

1. **Randomization:** Randomization of questions (selection of questions from a larger question bank and not just shuffling of order) is certainly the first feature that any MCQ based examination protocol must have. However, as our experiments show, the question bank size does not have to be as large as required in [9]. Now if the exam has only a few questions, all the questions can be of the same difficulty level. But for exams having more than 10 questions, it usually becomes necessary to have questions at multiple difficulty levels. A simple way to implement this could be to have a provision for dividing the questions into 3 levels, potentially easy (Level 1), intermediate (Level 2) and hard (Level 3). When students appear for the exam, they will be first shown questions from Level 1, then from Level 2 and finally from Level 3.
2. **Negative Marking:** Negative marking in MCQ based exams is important since it discourages students from doing guesswork. Typically, the weightage of negative marking for wrong answer is chosen to be 25% of the weightage of

positive marking for correct answer. Students should certainly be allowed to skip a question, in which case they got zero marks.

3. **Question Display:** If all the questions are displayed on the same page, it becomes easy for students to share the questions and copy the answers. Hence, we recommend displaying only one question at a time.
4. **Image Display:** If the exam questions are displayed as a text, it becomes easy for students to copy the questions and search for answers using various search engines and generative AI platforms like chatGPT (by OpenAI) or Bard (by Google). Hence, we recommend displaying each question as an image so that students cannot copy the question and search for answers online. There are certainly tools one can use to extract text from an image, but it is time taking and requires additional steps.
5. **No Back Button:** After a lot of deliberations with faculty members, it was decided that the questions displayed will not have a back button, which means that once the student has moved forward to the next question s/he cannot come back to the previous question. The student is allowed to skip a question to avoid negative marking, but in any case, there is no provision for going back to previous questions. This again has been recommended to prevent copying, but frankly speaking, it has been the only controversial feature of the protocol. Several students have expressed their displeasure with this feature since this makes the exam more difficult for them. In future experiments, we plan to introduce a back button and analyse how it changes the overall marks distribution.
6. **Bloom's Level and Topic:** In order to follow the OBE guidelines, while uploading the answer key on an online exam portal, the instructor should have the option of specifying the Bloom's Level and topic (or learning outcome) for each question as shown in Fig. 1, which makes it very easy for both the student and the teacher to figure out the difficulties faced by the students. This feedback can then be used by both the student and the teacher to focus on specific topics as needed.
7. **Detailed Feedback:** After the exam is over, both the student and the teacher should get to see a detailed analysis of the results as shown in Figs. 2 and 3, which helps in doing a comparative analysis and providing personalized feedback.

We evaluated several online assessment software platforms like Moodle[15], Canvas [16], Prairieconceptuarn [17], Google Form, etc and found some of these features available in these platforms, but were unable to find a single platform that provides all these features together in one place. Due to this reason, we decided to develop our own software portal, which is available for further experimentation and use by any course instructor [18]. We have also made sure that the process of setting up and conducting the exam is made as easy as possible for the teachers. We have conducted graded exams using this protocol at BML Munjal University and non-graded quizzes at Sitare University, and the results have been very promising. Informed consent was obtained from all participants and the study was approved and reviewed by competent authorities at BML Munjal University.

No	Answer	Bloom	Topic
1	D	Recall	LinearRegression
2	B	Understand	LinearRegression
3	A	Understand	LinearRegression
4	A	Recall	LinearRegression
5	C	Apply	LinearRegression
6	B	Understand	GradientDescent
7	A	Understand	GradientDescent
8	D	Understand	GradientDescent
9	D	Understand	GradientDescent
10	B	Understand	GradientDescent

Figure 1. Format of the CSV file to be uploaded to our online portal [18] while setting the exam. As can be seen, it allows the examiner to set the Bloom's Taxonomy level for each question as well as the topic (or learning outcome).

Q	Answer	Attempted	Skipped	Correct	%Correct	#A	#B	#C	#D	Topic	Bloom Level
1	D	24	1	12	48.0 %	1	1	10	12	POWER SERIES	RECALL
2	C	25	0	19	76.0 %	1	4	19	1	POWER SERIES	UNDERSTAND
3	A	24	1	13	52.0 %	13	4	3	4	POWER SERIES	APPLY
4	A	24	1	18	72.0 %	18	4	1	1	POWER SERIES	APPLY
5	D	21	4	6	24.0 %	7	5	3	6	POWER SERIES	APPLY
6	B	20	5	11	44.0 %	4	11	3	2	POWER SERIES	APPLY
7	A	23	2	20	80.0 %	20	2	0	1	POWER SERIES	APPLY
8	B	23	2	15	60.0 %	2	15	0	6	POWER SERIES	APPLY
9	C	24	1	22	88.0 %	2	0	22	0	POWER SERIES	APPLY
10	D	23	2	16	64.0 %	4	2	1	16	POWER SERIES	APPLY

Figure 2. Dummy example of feedback screen on our online portal [18], visible to the examiner. As can be seen, the examiner can conceptually see which questions were answered by a larger fraction of the class, and which questions were difficult for most students. This can help in designing the lecture material and future tutorial sessions. Examiner can share this with the students too for their own understanding and analysis of their performance.

Q	Submitted Answer	Correct Answer	Topic	Bloom Level
10	D	D	POWER SERIES	APPLY
11	Skipped	B	RIEMANN INTEGRAL	RECALL
12	A	B	MULTIPLE INTEGRALS	RECALL
13	Skipped	C	MULTIPLE INTEGRALS	APPLY
14	D	C	MULTIPLE INTEGRALS	APPLY
15	B	A	MULTIPLE INTEGRALS	APPLY
2	D	C	POWER SERIES	UNDERSTAND
3	B	A	POWER SERIES	APPLY
4	Skipped	A	POWER SERIES	APPLY
5	B	D	POWER SERIES	APPLY

Figure 3. Dummy example of Feedback screen on our online portal [18], visible to the students. This along with access to Fig. 2 can help the students in properly analysing his/her performance and figuring out ways of improvement.

IV. Results: Graded Exams

Our protocol was used to conduct four graded MCQ exams on CS related topics at BML Munjal University. For the sake of comparison, we also provide the marks distribution obtained for an offline MCQ exam conducted in one of the courses for which the online exam was done using our protocol. Table II shows the marks distribution obtained for four exams in online mode using our online protocol (Exams A, B, C and D) and one exam in offline mode (Exam E). Exam A was for a course on applications of Machine learning in engineering. Exams B and C were for one course (Human Computer Interactions). Similarly, Exams D and E were for one course (Data Structures and Algorithms). The subject name and other details of the exams have been hidden for the sake of privacy.

The detailed protocol followed in explained below:

1. **Exam A:** This exam was for a course on applications of Machine conceptuarning in engineering offered to around 30 third year students. The exam had a weightage of 5% in the overall marks, had 10 questions (all compulsory) and was conducted for 30 students in a room with one invigilator. Students were given +4 marks for correct answer, and -1 for wrong answer. Students did not experience any hassle during the exam, but were a little concerned due to lack of

back button.

2. **Exam B and C:** These exams were for a course on Human Computer Interactions offered to around 200 second year students. Exam B had a weightage of 5%, had 15 questions (all compulsory) and was conducted for 200 students across 4 different classrooms with one invigilator per room. Exam C had a weightage of 15%, had 45 questions (all compulsory) and was conducted for 200 students across 4 different classrooms with one invigilator per room. Students were given +4 marks for correct answer, and -1 for wrong answer. In both the quiz, students were informed prior that there will be negative markings and they cannot go back to previous questions. Sufficient time was given in both exams to answer all questions. In Exam B, students had some hassles and anxiety due to the newness of the platform, negative marking and lack of back button. However, in Exam C, these hassles were minimal since the students were already accustomed to the platform.
3. **Exam D and E:** These exams were for a course on Data Structures and Algorithms offered to around 300 first year students. Students were given +2 marks for correct answer, and zero for wrong answer. In this exam, we did not use negative marking to check if removing it leads to a positive shift in the marks distribution. Exam D had a weightage of 10%, has 15 questions randomly chosen from 23 questions divided into 3 levels. Students did not experience any hassle during the exam, but were a little concerned due to lack of back button. Exam D was conducted using our online protocol, and Exam E was conducted in offline mode (pen and paper). Exam E had a weightage of 10% and had 20 questions (all compulsory).

As can be clearly seen from Table II, conducting MCQ exams using our protocol helps in achieving a good marks distribution, which is similar to what one would obtain using regular offline tests or online tests with high degree of randomisation (see Table I). The distribution is maintained even when we did not use negative marking, as was the case for Exam D. The overall experience of faculty members was that our online protocol helps in reducing copying and guesswork, and so the marks distribution obtained gives a fair picture of the student performance.

Table II. Marks distribution obtained for four exams in online mode using our protocol (Exams A, B, C and D) and one exam in offline mode (Exam E). Exam A was for a course on applications of Machine Learning in engineering.Exams B and C were for one course (Human Computer Interactions). Similarly, Exams D and E were for one course (Data Structures and Algorithms). Out of these, Exam A-D were done using our protocol, and Exam E was done in offline mode. As can be clearly seen, conducting the exam on our protocol results in a marks distribution very similar to that obtained using offline mode.

Exam Name	Mean Marks	Standard Deviation
Exam A	43.25%	29.45%
Exam B	72.53%	19.56%
Exam C	49.2%	15.8%
Exam D	58.61%	18.67%
Exam E	53.5%	20.0%

V. Results: Non-graded Quizzes

Our online protocol was used to conduct weekly non-graded quizzes (i.e. marks obtained were not used for final grading) in two courses taught for 25 first year engineering students at Sitare University, and the results have been quite promising:

1. The non-graded quizzes provided useful feedback to the instructor, which was used to suitably modify the lecture contents and plan revision sessions.
2. Analysing the class performance through the OBE approach helped the instructor in figuring out the strengths and weaknesses of each student, and providing personalized feedback.
3. Students found this very useful for revision as well as to know where they need more practice, and have strongly recommended using this in general. An anonymous feedback was collected from the students to understand the efficacy of these weekly non-graded texts, and all the students who responded answered in the affirmative.
4. Implementation of our protocol through our own portal[18] made it very easy for the instructor to conduct weekly quizzes. Setting up the same quizzes on Google forms or other online portals would have been lot more time taking since there each question and each of the four options for each question has to be entered separately. On our online portal, the instructor just has to upload a consolidated PDF containing the questions and a single CSV file containing the answer key (see Fig. 1).

Overall the experience of following this assessment protocol for non-graded quizzes has been very effective, and we would strongly recommend all teachers to try it out.

VI. Discussion and Conclusion

In this paper, we have proposed a protocol for conducting MCQ based examinations, implemented it using our own online portal [18] and conducted controlled experiments in a few courses to validate its efficacy. The features of our proposed protocol clearly help in reducing guesswork and copying, as well as in analysing student performance in detail which helped in providing personalised feedback for further improvement. Over the next one year, it will be good to conduct more controlled experiments using our assessment protocol at other universities and colleges to collect more data to further establish the efficacy of this protocol as well as to make improvements.

One of the open questions is regarding the impact of negative marking on the marks distribution and student performance. It is generally believed that negative marking reduces guesswork, but high negative marking can deter

students from even attempting some questions they could have otherwise answered correctly. Also, as mentioned earlier, one of the features of our protocol that caused a lot of anxiety for students is the absence of a back button. We also plan to do some controlled studies with the back button present. And a broader question is regarding the applicability of our proposed protocol in other educational settings. Answering all these questions will require further work, which we will be taking up in the near future.

The data collected by using our OBE based protocol can also help in designing personalized AI assistant for school and college education. A vast majority of students require personalized attention in order to learn various subjects properly and perform well in exams. However, given our large population size, providing personalized attention is not practically possible. This is where Artificial Intelligence (AI) can come in handy by acting as a personal assistant that students can interact with to learn effectively. The idea is not to replace human teachers, but augment them to get the best out of both. Human teachers are needed to teach, while AI can help in effective assessments and personalized feedback for further growth. There are a lot of startups working on developing such a personalized AI assistant, but their efficacy is not clear since neither the algorithm nor the student data is available in public domain. It is very important to build open source AI tools for personalization of education due to its vast impact and immense social importance. And building AI tools requires data, which such MCQ based examinations based on OBE can provide.

We will be glad to have faculty members from other universities, colleges and schools to be a part of this experiment. This can lead to a significant improvements of our automated assessment and feedback systems, eventually leading to an improvement in the overall learning outcomes in large classes.

Acknowledgments

We would like to thank Prof. Shyam Menon (Vice Chancellor, BML Munjal University), Prof. Maneek Kumar (Dean, School of Engineering and Technology, BML Munjal University), Prof. Anirban Chakraborti (Dean Research, BML Munjal University), Prof. Jaskiran Arora (Dean Education Quality, BML Munjal University) and Prof. Suneet Soni (Controller of Examinations, BML Munjal University) for providing valuable feedback and supporting this experiment.

References

1. ^*URL : https://www.forbes.com/quotes/4638/ (last accessed : May 09, 2023)*
2. ^*M. Nunez-Canal, M. M. Obesso, and C. A. Pérez-Rivero. New challenges in higher education: A study of the digital competence of educators in Covid times. Technological Forecasting and Social Change 174, 121270 (2022)*
3. ^*J. Grodotzki, S. Upadhya, and A. E. Tekkaya. Engineering education amid a global pandemic. Advances in Industrial and Manufacturing Engineering 3, 100058 (2021)*
4. ^*J. L. Burnette, C. L. Hoyt, V. M. Russell, B. Lawson, C. S. Dweck, and E. Finkel. A Growth Mind-Set Intervention Improves Interest but Not Academic Performance in the Field of Computer Science. Social Psychological and*

Personality Science 11, 107 (2020)

5. ^*URL : https://www.thehindu.com/news/cities/bangalore/computer-science-continues-to-dominate-engineering-admissions-this-year-too/article66027135.ece (last accessed : May 09, 2023)*
6. ^*C. Mulryan-Kyne. Teaching large classes at college and university level: challenges and opportunities. Teaching in Higher Education 15, 175 (2010)*
7. ^*M. Henderson, T. Ryan, and M. Phillips. The challenges of feedback in higher education. Assessment and Evaluation in Higher Education 44, 1237 (2019)*
8. ^*URL : https://www.indiatoday.in/education-today/news/story/over-80-indian-engineers-are-unemployable-lack-new-age-technology-skills-report-1483222-2019-03-21 (last accessed : May 09, 2023)*
9. a, b, c, d*M. R. Santos, V. Richman, and J. Jiang. Online Teaching: A Study for the Effectiveness of Randomized Exams. Journal of Instructional Pedagogies, 22 (2019).*
10. ^*W. Spady. Outcome-Based Education: Critical Issues and Answers. Arlington Virginia: American Association of School Administrators (1994)*
11. ^*S. M. Allais. Education service delivery: The disastrous case of outcomes-based qualifications frameworks. Progress in Development Studies 7, 65 (2007)*
12. ^*URL : https://www.ieagreements.org/accords/washington/signatories*
13. ^*URL : https://www.education.gov.in/sites/upload_files/mhrd/files/*
14. ^*B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, D. R. Krathwohl. Taxonomy of educational objectives: The classification of educational goals. Vol. Handbook I: Cognitive domain. New York: David McKay Company (1956)*
15. ^*URL : https://moodle.org/*
16. ^*URL : https://www.instructure.com/lms-learning-management-system*
17. ^*URL : https://www.prairielearn.com/*
18. a, b, c, d, e, f*URL : https://www.bekushal.com*