A First-order Action Research Study to Uncover Students' Conceptual Gaps in an Online Statistics Course using Extended Matching Questions

Ishaan TANEJA[a*], Prajish PRASAD[b], Jayakrishnan WARRIEM[a]
[a]*Indian Institute of Technology Madras, India*
[b]*Department of Computer Science, FLAME University, India*
*ishaan.taneja.stats@gmail.com

Abstract: The COVID-19 pandemic has forced institutions worldwide to accelerate their practices on digital transformation, especially in the use of Online Learning. The rapid pace of this transition has forced instructors also to adapt quickly to the online instructional design and in turn reduced opportunities of systematic action research within such a setting. In this paper, we present the action research study by an online instructor to tackle the problem of understanding conceptual gaps among learners. The instructor adapted an existing solution widely used in medical learning, called extended matching items, to design 'Extended Matching Question (EMQ)' for large scale online students (N=1525) of an introductory Statistics course. The instructor modified the structure of this assessment by allowing learners to first identify the concepts involved to solve a particular question and then use those concepts to reach the final numeric answer in the topic 'Permutations and Combinations'. Comparisons of this assessment were made with regular assessments (multiple choice or numerical answers) on the efficacy to make students apply their conceptual understanding for solving the problems. The analysis also enabled the instructor to identify the exact conceptual points where the students lag and showed strong association between identification of concepts and getting the final numeric answer correct. In addition, there was a significant increase in the average score of the test takers in the EMQ assessment compared to that of regular assessment. The study also presents pointers to universities involved in hybrid learning regarding the need for setting up research-practice partnerships within their institutions to tackle their own problems of practice.

Keywords: Action research, Research Practice Partnership, Extended Matching Questions, Statistics, Online Assessment

1. Introduction

The COVID-19 pandemic has forced institutions worldwide to accelerate their practices on digital transformation, especially in the use of Online Learning. Researchers have already pointed out the significant differences between carefully planned online instructional design and a rapid transition to online teaching (Hodges et al., 2020) that results in widespread rejection of online teaching methodologies adopted during the COVID-19 pandemic. Two practical difficulties exist in the systematic design of online instructional processes in such a setting - (i) inability for teacher professional development programmes to match the rapid pace at which institutions are going online (Byrne et.al., 2021), (ii) Limitations of the technology infrastructure in accommodating pedagogical designs. However, this also presents a unique opportunity for researchers to trace the evolution of online instructional practices, especially that of teacher action research in an online setting.

The primary goals of action research studies are to improve the practice with a subordinate goal to produce knowledge that may then be utilized for improving practice (Elliot, 1998). Posch (2019) explained these two goals as two different interests - developmental interest (What would I like to

develop or change?) and research or cognitive interest (What would I like to find out and understand?). The action research studies catering to the former interest are called first-order action research (Hanfstingl & Pflaum, 2022) and will typically involve identification of a problem, suggest possible solutions, implement the solutions and collect information to see effectiveness of their solutions (Altrichter et. al., 2002). The context of large-scale hybrid online learning (similar to Massive Open Online Courses) brings in another complexity to the process as there is a limitation in conducting qualitative studies on specific learning instances across the duration of the course. In this paper, we present a first-order action research study that explains adaptation of an assessment strategy to uncover students' conceptual gaps in an online Statistics course offered at scale.

2. Identifying a problem of practice in the operating context

2.1 Description of the Statistics course in the hybrid online programme

The context of this action research is a large-scale introductory Statistics course which is part of a hybrid Undergraduate degree program in Programming and Data Science at a University in India. The primary mode of teaching-learning within the program is Online (using LMS and Video conferencing services), and there is a mix of online and in-person assessments (conducted simultaneously across various assessment centers across the country). The course is of 12-week duration with weekly online assessments and the in-person assessments occur after the fourth week. The scores in the weekly assessment determine the eligibility of the student to write the in-person assessment. The scores in the in-person assessment contribute to the overall grade of the learners and determine their eligibility to pursue the longer degree programme.

The 12 weeks of the introductory Statistics course is split into three parts: (i) first part (i.e. Week-1 to Week-4) covers the concepts relevant to summarization and visualization of the different types of the data, (ii) second part (i.e. Week-5 to Week-8) of the coursework covers the skill of estimating the chance by applying laws of probability, (iii) the last part (i.e. Week-9 to Week-12) covers the understanding and properties of Random variables along with some probability distributions. The concepts and skills are aimed at helping students proceed to higher courses in Statistics and Machine learning that are required for data scientists.

The primary form of instruction were video lectures which were recorded beforehand by course faculty, and were released every week. The course also had instructors who were responsible for day-to-day orchestration of the course (The first author was an instructor of this course). The primary responsibilities of instructors were to create assessments and interact with students. Interactions with students happened synchronously through live sessions, and asynchronously through discussion forums. These synchronous and asynchronous interactions were useful in addressing students' concerns and difficulties they had with regard to contents of the course.

2.2 Description of the Course Learning Management System

In the course portal, students get access to all the content of coursework in which they are enrolled. The content in the portal is updated as per the release date of weekly contents. The front view of the portal is shown in Figure 1 below.

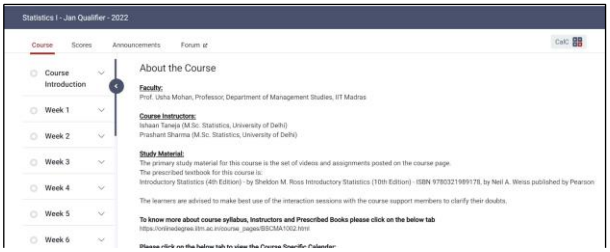

Figure 1. Student view of the portal

By clicking on the Week 'X', the student can access the content of X^{th} week (an example shown in Figure 2), which consists of lecture videos each followed by assessments to test the students' understanding of that specific lecture. At the end of all the video lectures, there are some tutorial videos recorded by the instructors followed by the practice assignments and the graded assignment (formative assessments) which test students' knowledge on the week's content.

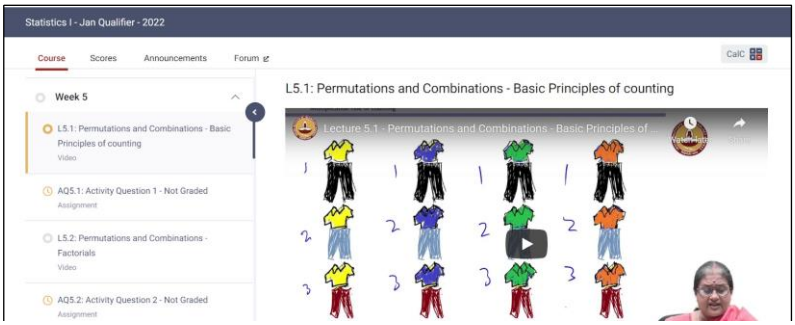

Figure 2. Weekly content view on the portal

2.3 Assessments used in the statistics course

In the Statistics-1 coursework, we test students using a variety of assessments, i.e. Activity questions, Practice assignments, Graded assignments and Quizzes. The 'Activity Questions' are to be attempted after going through each lecture video, in which the students are tested only on the example and conceptual statements used in the lecture video, to make sure that they understood what's been explained in the respective lecture video.

The Practice assignments and Graded assignments contain numerical answer tests (NAT), multiple choice questions (MCQs) and multiple select questions (MSQs). Both Practice and Graded assignments are similar in terms of pattern (approximately 50 % NAT, and remaining 50% MCQ and MSQ) and difficulty. Both of these assessments test the ability of students to apply the concepts taught in the respective week's lecture videos. There's a deadline to submit the Graded assignment, for each week students have a week of time to go through weekly content and solve this assessment. A student can also make multiple submissions to it, but only the last one will be considered for the scoring.

2.4 Limitations in assessments for measuring student learning of statistics concepts

As mentioned in Section 2.3, the type of assessment questions used to test students' learning from the content provided consists of MCQ/MSQ and NAT questions. On examining the assessment data from the LMS, the instructors observed that many students fail to get correct answers for the NAT questions. Further analysis of individual answers revealed that most often students lose out on the marks due to calculation errors. Hence, students might know the concepts and steps required to solve the answer, but NAT type of questions fail to measure their conceptual understanding.

Two of the most effective ways to overcome the drawbacks of such assessments are: a) use questions from Statistics Concept Inventory (SCI), or b) to replace them with subjective assessments. SCI consist of MCQs having options that correspond with common misconceptions that students may have in a particular topic (Stone et. al., 2003). An understanding of the misconceptions present among the students can further help the instructor plan for redress strategies. However, the instructor will not be able to track the exact position where the student develops this misconception by just using SCI as there could be intermediate steps in solving the problem. Similarly, it is not feasible to conduct and evaluate subjective assessments weekly at scale. Hence, we needed a method which can address this gap of conceptual testing of students at scale.

In order to address this problem, we analyzed existing literature for assessment strategies which test conceptual understanding of students. We found that Extended Matching Questions (EMQ) assessment

type turns out to be one of the potential solutions that not only overcome the limitations of subjective assessments but is also a good measure of conceptual knowledge of the test takers. EMQs have been adopted in multiple contexts like medical and healthcare education, numeracy assessments (Wood, 2003; Cann, 2005). However, we have not seen any reported instances of use of EMQs in teaching-learning of Statistics, especially in a fully online mode of delivery with a diverse learner population (in age and ability). In the next section, we explain the details of this solution, and how we incorporated it into a fully online course setting.

3. Design of Solution - Extended Matching Questions

3.1 Description of Extended Matching Questions (EMQ)

Extended Matching Questions (EMQ), have long been prevalent in medical education, and are similar to multiple choice questions but with one key difference, that they test knowledge in a far more applied, in-depth, sense (Wood, 2003). EMQ consists of a list of more than 8 options along with a few number of questions that will have some of these options as a valid intermediate solution (as a diagnosis or root cause) or final solution. The design of questions and options are roughly similar in structure and content, to test the in-depth knowledge.

A diagrammatic structure for an EMQ is given in Figure 3. In the designing of a list of options for EMQ assessment, one needs to ensure that each question must have at least one best/correct answer from among the series of options given. One option can be an answer to more than one question, or may not even be an answer to any of the questions. This will ensure that the probability of getting the correct answer by chance remains reasonably low. For example, from Figure 3 we see that for Question-1, option-1 is the only correct option in the list of 7 options, which implies that the chances of getting it completely correct by guesswork is approximately 14 percent. Similarly, for Question-2, since three options out of the seven are correct, the total number of ways to answer the question is 35 ($^7C_3$). Hence, chances of getting it completely correct by luck is approximately 3 percent. If this was an MCQ with a list of 4 options, the chances of getting it correct would be approximately 25 percent.

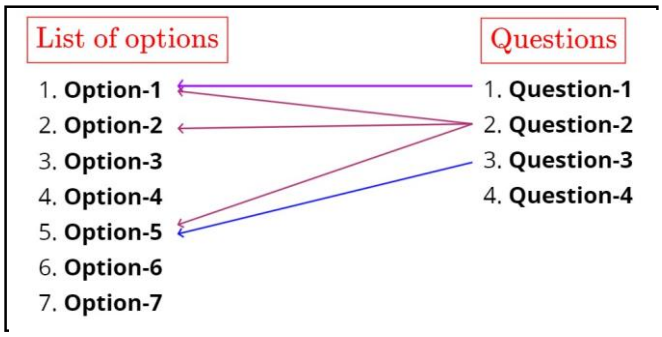

Figure 3. Diagrammatic Structure for EMQ

EMQ has often been used in medical education and other healthcare subject areas for diagnostic reasoning (Case & Swanson, 1993; Wood, 2003; Samuels, 2006; Vuma S, Sa B, 2017). It is evident from research that students attempting this test item format have a greater chance of answering incorrectly if they cannot synthesize and apply their knowledge (Case & Swanson, 1993). Furthermore, it also ensures that the probability of getting the correct answer by chance remains reasonably low. One of the major aims while designing an Extended Matching Question is to condense the list of options, as we don't want a list of 40-50 options for a 10 questions test. Hence, options were designed in such a way that if a student makes a conceptual mistake and gets some answer, say, 'X', then that answer will also be one of the options. Also, that option might be the solution for another question in the list.

We believe that EMQ have several advantages over an MCQ and NAT assessment. In an MCQ assessment with four options, the chances of getting any question correct is 25% by a guess (in an ideal

case). However, in case of an EMQ assessment, this probability reduces significantly. Also, in case of an NAT, there is a good chance that the person taking the exam has good conceptual knowledge in the topic tested but he/she may have got the answer wrong due to some calculation error. There is no way to assess whether the error is due to a lack of conceptual knowledge or due to a calculation error. EMQ assessments can help address this and can indicate to the course instructor which topics a student is lacking in and needs more attention.

3.2 Implementation of EMQ in the online statistics course

The basic understanding of 'Permutations and Combinations' plays an important role in all the upcoming weeks. Therefore, it is necessary to identify as an instructor where the students are lagging and more specifically the exact point of conceptual gap. We implemented EMQs in the Week-5 content of Introductory Statistics course that covered the content - 'Permutations and Combinations'. The idea was to split the graded assignment of the week into two parts. The first part will be the regular assignment (NAT and MCQ) while the second part will be EMQ assessment. Each question in the first part will have a corresponding implementation in the EMQ part. Example questions of regular assessment types are shown below in Figure 4 and Figure 5.

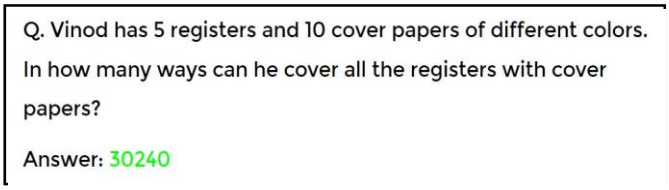

Figure 4. NAT variation of a question

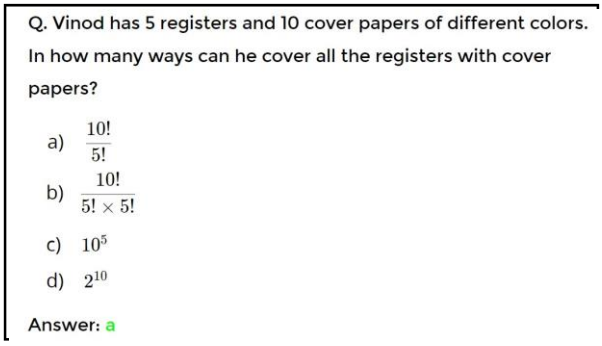

Figure 5. MCQ variation of a question

To ensure that drawbacks of MCQ and NAT are handled, we designed the corresponding EMQ variation as shown below in Figures 6 and 7.

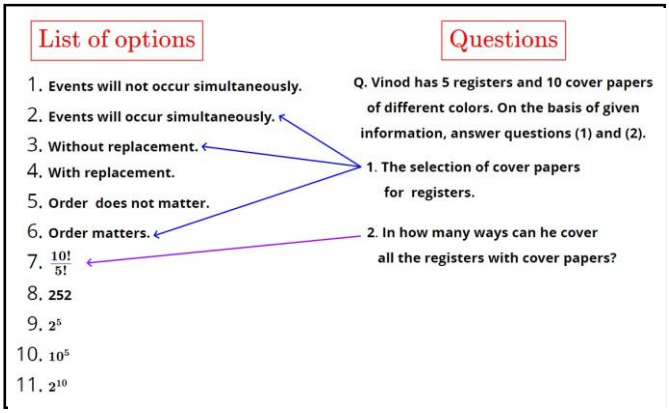

Figure 6. Diagrammatic representation of EMQ variation of a question

Each EMQ question is divided into two subparts, where the first part asks about the conceptual steps required to solve the question, while the second part requires solving and reaching the final solution by applying the conceptual steps selected in the first part of the question. For example, as shown in Figure 7, the first part of questions requires the test takers not to solve the question but just to identify the steps that are necessary to compute the final answer, while the second part requires the computational part to be done. In this way, we can get an idea if students need special attention to this topic, then what to focus more on. Since, the set of questions are from a specific topic, there is a high chance that the conceptual steps used to solve any numeric question relevant to that topic may be the same. As shown in Figure 6, there are six options which are relevant to step identification part of the question from the section 'Permutations and Combinations'. All questions in this assignment will be required to identify steps from the given list of six options only.

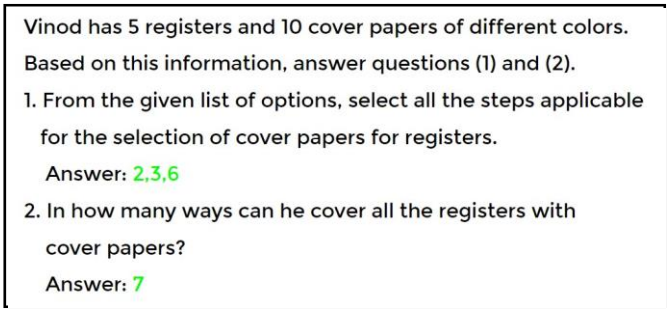

Figure 7. EMQ variation of a question

Before giving such assessments as graded assignments, we also needed to make sure that the test takers are aware of such types of questions and know how to answer them. Hence, we gave them practice EMQ assignments along with the Week-2 (Describing Categorical Data) and Week-4 (Association between two variables) content of the Statistics-1 course. There were also live video conferencing sessions explaining the motive of such assessments and details of the steps to be followed to submit the answers.

3.3 Extending the functionalities of course portal to incorporate EMQ assessments

The course portal was designed to conduct the regular assessments such as MCQ/MSQ, NAT and programming assessments (in input-output testing mode and functional testing mode) only. To enable EMQs, we made use of the functional testing mode of the programming assessments that allowed evaluations by custom scripts written by the course instructors. The functionality allows students to upload their answers to the EMQ via a '.txt' file which is then evaluated using the script provided by the instructors (see Figures 8 and 9). The use of script also allowed the instructors to provide formative feedback to the students.

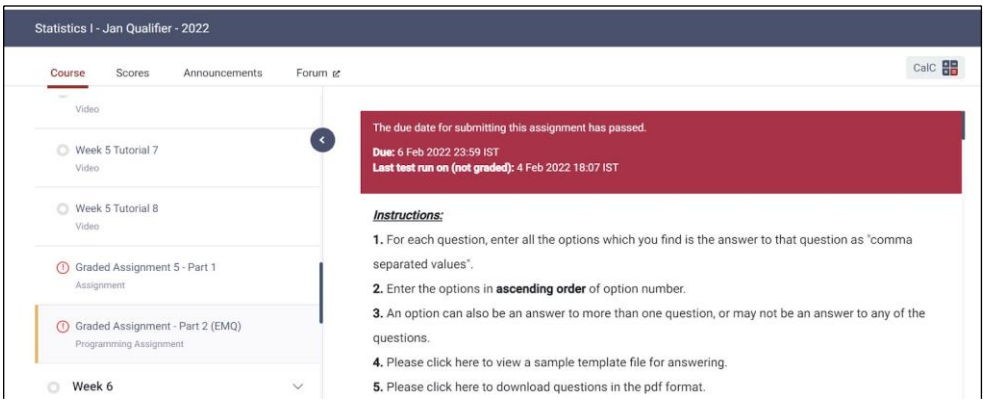

Figure 8. Student view of EMQ assessment on the portal

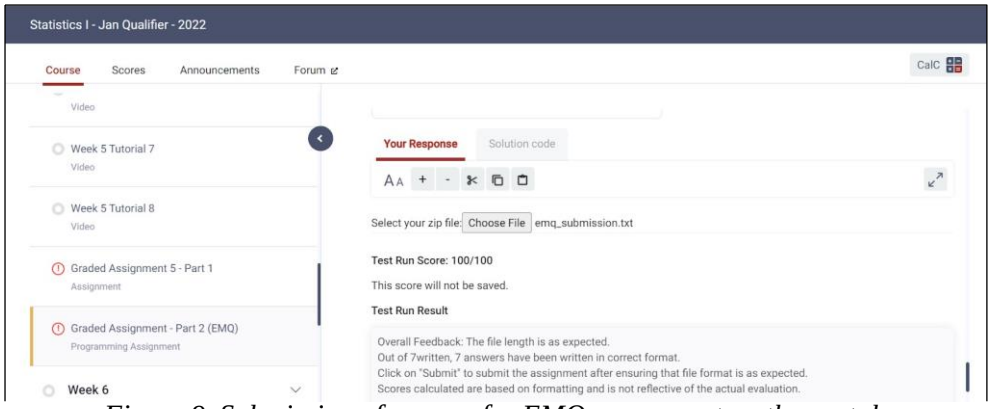

Figure 9. Submission of answer for EMQ assessment on the portal

4. Effectiveness of the EMQ Solution in addressing limitations of previous assessments

4.1 Students' Performance in EMQ Assessments

In the EMQ assessment, we received responses from 1109 students of which 1004 agreed for the use of their inputs and aggregate scores to be used for research purposes. While in regular assessment (MCQ and NAT), 1316 of 1525 students agreed for the same.

Firstly, we analyzed the inputs of test takers to step-wise questions in EMQ assessment. Secondly, we considered those students who were able to identify at least one of the steps correctly and analyzed if it affects the percent of correct final answer to the problem, i.e. second sub-part of the question in which the test takers have to identify the final numeric answer to the problem. We also computed the result for regular assessment, of all those students who attempted the question, and tabulated it in Table 1.

Table 1. *Comparison of EMQ and Regular assessment*

Question Number (Type of Question)	EMQ (At least one step correct)	EMQ (All step correct)	% of students correct in NAT/MCQ
Question - 1 (MCQ)	93.2 %	98.4 %	66 %
Question - 2 (MCQ)	48.4 %	78.1 %	53.8%
Question - 3 (NAT)	48.8 %	36.4 %	41.4 %
Question - 4 (NAT)	93.4 %	95.1 %	77.9 %
Question - 5 (NAT)	91.2 %	94.7 %	57.4 %

The first two columns of the table correspond to the EMQ assessment result, and the last column is relevant to the result of regular assessment questions. The first column of the Table represents the percent of students who got the final numeric answer correct given that they identified at least one of the steps correctly, and the second column represents the percent of students who got the final numeric answer correct given that they identified all the steps correctly. While the third column represents the percentage of students who attempted the regular assessment question and got it correct.

From the tabulated data, it can be concluded that if the test takers identify all the steps involved in solving the problem, then the chances of getting the correct final numeric answer increases significantly and even if at least one step is identified correctly the chances of applying conceptual knowledge to get the final answer correct is more than that of regular assessment question. The pattern noticed for Question - 2, is a bit different from others and on further analysis we identified that the test takers struggled in identifying the step if order matters or not, which helps the faculty and instructors to focus more on such similar questions and give students a better understanding of it, i.e. where the test takers are facing problem. Another strange pattern is noticed in Question - 3, where the correct percent to final answer remains consistent irrespective of assessment type and steps identification in EMQ assessment. It is because this question involves the concept from Week-1 content as well, i.e. we connected the topic

'Scale of measurements' with the 'Permutations and Combinations', which students may not be able to recall while solving and hence we saw a big dip in the percentage of students who got it right. Therefore, from this, it can be concluded that without any intermediate steps related to Week 1 and 2 in the EMQ option, the correct percentage falls.

The aggregate scores of test takers in both the assessments are tabulated in Table 2. In the regular assessment, only the five questions which were similar in concept to EMQ assessment questions were considered for this.

Table 2. *Average scores of EMQ and Regular assessment*

	EMQ (%)	NAT/MCQ (%)
Overall	62.09 %	57.42 %
Students who attempted at least one question	75.11 % (830 students)	57.50 % (1318 students)

We performed t-test for independent means to test if the average score in EMQ assessment is significantly greater than that of regular assessment score of those students who attempted at least one of the questions in the assessment. The t-test gives a p-value of 0.000 (< 0.05). Hence, it can be concluded that the scores of EMQ assessment is significantly greater than that of regular assessment at 5 percent level of significance.

Thus, all this implies that EMQ assessment is not only a better way to test the students but is equally helpful for faculty and instructors. As it makes students apply their conceptual knowledge to solve the question and not by the rote memorization of when to do multiplication/addition, which results in improved scores of the test takers. On the other hand, the faculty and instructors are able to identify the weak spots of students in 'Permutations and Combinations' in certain questions and can accordingly work on the contents and interaction sessions with students.

4.2 Students' Perspectives on EMQ

We sent a feedback form to students to get their feedback on EMQ after the EMQ type practice assignments in Week-2 and Week-4 content of the Statistics-1 coursework. The main aim of the feedback form was to get an idea of students' view towards EMQ and any trouble they were facing in answering the EMQ assessments. In the Week-2 feedback form, we had 97 responses, in which approximately 60 percent of test takers found the regular assessment difficult compared to the EMQ assessment, and more than 70 percent of test takers wanted more such assessments in the upcoming weeks. But the majority of the audience wanted the way of answering to be more convenient, as they have to repeatedly scroll up and down to see the list of options and read the question.

To resolve this issue rapidly before the Week-4 content releases, we provided the PDF of questions along with the assessment too in order to avoid the struggle of test takers to repeatedly scroll up and down. But the way to submit answers, i.e. uploading ".txt" file is still a tedious task from students perspective and could be taken care of in future updates of the portal where they will be provided a textbox within the portal to input the answers.

After the Week-4 EMQ assessment, we received 27 responses to the feedback form, in which the majority of students (more than 55 percent) believed that even though EMQ assessment takes more time, it requires more understanding of the underlying concepts as compared to other types of questions. In addition, after these assessments the test takers believe that *"EMQ not only tests their understanding of concepts but also avoids the tedious calculations which helps in better decision-making."*

5. Conclusion

In this paper, we described how instructors adapted an assessment strategy to uncover students' conceptual gaps in an online Statistics course offered at scale. Based on student difficulties encountered in the course, instructors adapted 'Extended Matching Question (EMQ)' which has been primarily used in medical education, to identify the gap in students' conceptual understanding in the topic 'Permutations and Combinations'. This assessment type overcomes the limitations of the regular assessments, i.e. NAT and MCQ, as it helps students think and apply their knowledge conceptually to reach the final numeric answer rather than by the memorization of steps or guesswork. In addition, it also helps instructors capture the exact conceptual point where students need more attention, which helps in the improvement of contents in the coursework.

We evaluated the effectiveness of EMQ assessments by measuring student performance and taking perspectives of students and course instructors. Students attempted both assessments, i.e. regular type and EMQ type, each based on the similar concepts. Comparison of EMQ and regular NAT/MCQ assessments show that there is a significant increase in the scores of students in the EMQ type assessment as compared to that of regular assessment. Based on the analysis of students' answers to the EMQ type assessment, it is found that as the ability to apply the conceptual understanding of concepts increases, there are very high chances of getting the final numeric answer correct. Instructors also realized that students were struggling with certain concepts and question types, and this helped them plan their tutorial sessions with the students in the upcoming weeks.

This study is also a first order action research where the focus has been on addressing the development interest of the action researcher (Byrne et. al., 2021). We observe that the action researcher has systematically studied the problem at hand, designed and implemented a solution, collected data related to the implementation and analyzed the outcome of the implementation (Altrichter et. al., 2002). We observe that the online setting makes it easier for instructors to collect data related to their interventions and thus making quantitative methods of research more amenable. Further the roles of research mentors (in this case, second and third authors) in the research-practice partnership is critical in providing a meaningful structure for the action research. In the current study, the first author had consistent meetings with the research mentors to identify systematic ways of validating the findings and further structuring the write-up for the entire study. Universities involved in hybrid learning can invest in creating such research-practice partnerships as they provide these institutions with solutions to immediate problems of practice as well as larger research problems related to second-order that they can invest in for a longer term.

There are two limitations to our study. The first limitation is on the difference in the count of students who attempted both the assignments. We observe that there's a small portion of students who have not attempted the EMQ assignment or whose inputs were not stored due to formatting error while submission, but the size of the audience is large enough to overcome this effect while analyzing by aggregating the results of each of the two assignments. The second limitation is that, it could also be possible that the few may have attempted it just to get an idea of the assessment for future reference, without going through the content as they also have an option to opt out from the course end assessments. We intend to address these limitations by conducting longitudinal studies by providing EMQ assessments across weeks. This can give us a better understanding of the benefits of EMQ and how students' conceptual understanding is improving as the course progresses.

We believe that Extended Matching Questions can be adapted to other topics in Statistics and Maths as well. As part of our future research, we intend to use EMQs in other relevant Statistics and Math courses in the programme in a more longitudinal format. We believe that the results of this study will generate sufficient interest among instructors to move towards the higher levels of scholarships in teaching and learning.

References

Altrichter, H., Kemmis, S., McTaggart, R., & Zuber-Skerritt, O. (2002). The concept of action research. *The Learning Organization*, 9(3), 125-131

Byrne, V. L., Hogan, E., Dhingra, N., Anthony, M., & Gannon, C. (2021). An exploratory study of how novice instructors pivot to online assessments strategies. *Distance Education*, *42*(2), 184-199.

Cann, A. J. (2005). Extended matching sets questions for online numeracy assessments: a case study. Assessment & Evaluation in Higher Education, 30(6), 633-640.

Case, S. M., & Swanson, D. B. (1993). Extended-matching items: A practical alternative to free-response questions. *Teaching and Learning in Medicine: An International Journal*, *5*(2), 107-115.

Elliot, J. (1991). Action research for educational change. McGraw-Hill Education (UK).

Hanfstingl, B., & Pflaum, M. (2022). Continuing professional development designed as second-order action research: outcomes and lessons learned. *Educational Action Research*, *30*(2), 223-242.

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). The difference between emergency remote teaching and online learning. *EDUCAUSE Review*, Accessible at: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *Journal of Online Learning and Teaching*, 8(3), 198-208.

Posch, P. (2019). Action research–conceptual distinctions and confronting the theory–practice divide in Lesson and Learning Studies. Educational Action Research, 27(4), 496-510.

Samuels, A. (2006). Extended matching questions and the Royal Australian and New Zealand College of Psychiatrists written examination: an overview. *Australasian Psychiatry*, *14*(1), 63-66.

Stone, A., Allen, K., Rhoads, T. R., Murphy, T. J., Shehab, R. L., & Saha, C. (2003, November). The statistics concept inventory: A pilot study. In 33rd Annual Frontiers in Education, 2003. FIE 2003. (Vol. 1, pp. T3D-1). IEEE.

Vuma, S., & Sa, B. (2017). A descriptive analysis of extended matching questions among third year medical students. *Int J Res Med Sci*, *5*(5), 1913-1920.

Wood, E. J. (2003). What are Extended Matching Sets Questions?, *Bioscience Education*, 1:1, 1-8.