



An Exploratory Study on the Effect of Random Question Sequencing in Computer-Based Testing (CBT) on Student Performance

¹Abdulahakeem, I. and ²Sabiu, L. T.

¹Department of Computer Science, Federal University Birnin Kebbi

²Management Information System Directorate, Federal University Birnin Kebbi

ABSTRACT

The Computer-based test has become a handy tool in delivering examinations in our institutions of learning, and it has proven to reduce the time taken in marking and releasing results, improve teacher's productivity, and reduce malpractice during the tests. Various techniques to further reduce malpractice while taking CBT examinations have been proposed, and one of them is to randomize the questions such that adjacent students will not have the same question sequence, and thus, reduce copying and cheating during the exams. The research hypothesized that randomizing questions and/or options could pose tests and other emotional anxieties to students; most especially on those given difficult questions first compared to those with easier questions first - which in turn could greatly affect the performance of the students. Two hundred and sixty-four (264) participants that took a course on Introduction to Computing were recruited and then divided them into three groups. The first group took CBT exams with the same sequencing of questions and the second group took the exams with random sequencing of questions with difficult questions first and the last group took the exams with random sequencing with easy questions first. The study found that the students in the first group performance are higher than those in the latter groups. The research recommends that instructors should balance the nature of sequencing the randomized CBT questions to reduce the anxieties that affect the students when considering giving examinations online.

Keywords: *Computer Based Test, Random Sequencing, random sequencing, question difficulty, examination malpractice*

INTRODUCTION

Institutions of learning are increasingly adapting and adopting the computer-based test as another method of assessing students. This is mostly due to it being simple, cheap and speedy result computation and compilation. Anxiety due to examinations or using computers has been extensively studied and shown to have an impact on student performance, and will refer the reader to (Marks, 2008) for an extensive review. This has largely been attributed to the lack of complete control that students have on computers than papers - such as reading all the questions and jumping between questions (Ricketts and Wilks, 2002). Another factor includes strange or unfamiliar software/ hardware that the students find to be new to what they are normally used to. While trying to reduce the impact these changes have on students, malpractice is another challenge that needs to be addressed. Students tend to cheat by copying from other during CBT examinations. Some interesting methods and techniques are employed to minimize this too; such as setting some distance between students, using screen front polarizer, privacy filters (Joe and Charles 1999). The most common method is to randomize the sequence of questions (some software even go further and randomize the options too) (Pain and Heron, 2003; Wang *et al.*, 2004). It is quite effective in preventing cheating from looking over other student's computer screens. It is also cheap, as it needs no additional hardware. The randomization of questions though comes with a price, considered to be very high. It creates some test anxieties to students due to some of them having

simple questions coming first and difficult ones last, or vice versa. It also adds a cognitive burden to some disadvantaged students. This violates the APA guidelines that require fairness to all test-takers (APA, 2008). This study inspected and measured the impact of randomization and test anxiety on the student's performance.

METHODS

This section will detail out the methods used in carrying out the study. We will explain how the participants of the study were selected, how questions were categorized and grouped which serves as the basis for the study. We then explained how the examination was carried out and recorded.

Participants

Two hundred and sixty-four (264) students were enrolled in Year 1 *Introduction to Computer Science Course* in exchange for some credits. There were 162 males and 102 females in the participant pool. 100% percent of the participants have taken CBT examination before (from JAMB¹ or other sources), and thus familiar with using computers and taking an online test. This helps in removing anxiety from computer usage as the study wants to concentrate on test anxiety from randomizing questions and its effect on the performance of the test-takers.

¹<https://www.jamb.org.ng/>

Categorizing Question

Categorizing questions based on their difficulty is an important aspect of setting up tests. This helps the examiner to identify which of the questions need to be retained, edited or discarded (Alessi and Trollip, 2001). According to Reise and Henson (2003), the difficulty index of a question is computed as the ratio between the number of correct responses for a question by the number of attempts (both correct and incorrect) made to answer that question, which can be expressed as a percentage. When a question has a high

difficulty index, it means it is expected to be difficult, and that with low difficulty index is expected to be easy. To do this, Seven (7) experts were enrolled in the module (lecturers) to tag and categorize the questions from our question bank accordingly. Another 41 average students that took the module in the past year (those with at least a C and at most an A) were enrolled to answer all the questions in the data bank while their attempts were logged on each question. Then this information was used to compute the difficulty index of all the questions and store them in the data bank.

Table 1: Categorizing Questions based on their difficulty index

Questions	Experts	Students
Easy Questions	79%	73%
Difficult Questions	21%	27%

Table 1 shows the result of categorizing questions according to easy or difficult based on their difficulty index. Experts and student results showed synergy between the results of the categorization and ranking. While experts ranked 79% of the questions in our database to be easy, students found only 73% to be so and students marked 27% of the questions to be difficult, while experts consider only 21% to be difficult.

Examination

The students were categorized into three groups by sorting both genders by their admission numbers and then dividing them into three groups. Each group has 54 males and 34 females. The first group (subsequently called Group A or Normal Group) served as the control group. Their questions were generated by alternating between simple questions and difficult questions. The second group (subsequently called Group B or Easy Group) had their questions generated with all the simple questions first, then difficult questions to the end. And finally, the last group (subsequently called Group C or the Difficult Group) had their questions generated with difficult questions first, then easy questions at the end.

Each group used a different examination center with all centers starting at the same time to avoid sharing information about the exams between the participants. The exam contains topics that were covered with the students in the classroom and are all multiple-choice questions. The

duration of the test is 30 minutes to complete the examination. In-house made CBT software developed for the University to deliver online examination was employed and made some modifications to allow the randomization effect. Students are allowed to answer the questions in any order they desire. During the exam, each student was given 15 multiple choice questions to answer in 30 minutes (an average of 2 minutes per question).

Analysis

The study uses descriptive and inferential statistical methods of linear regression to analyze the results on SPSS². The research employs a special case of ANOVA³ to analyze the gender characteristics of the population, and define and compare the performances of the three groups. The research also considered the ethical implications of the study by ensuring that participants agree and consent to be part of the study and can withdraw from the study at any time.

RESULTS AND DISCUSSION

The first result is the average performance of students of the three groups compared to their gender. The research wanted to investigate if the nature of questions has more impact on one gender compared to another. From Figure1, it is clear that there is no significant difference in the performance of males as compared to females in all the groups.

² <http://www.spss.com.hk/>

³ https://en.wikipedia.org/wiki/Analysis_of_variance

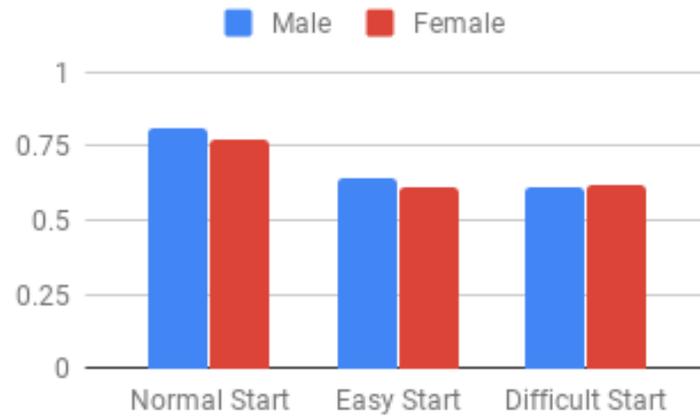


Figure 1: The performance of participants based on gender across the three groups

This is because the students underwent the same training and female students tend to perform as good as their male counterparts. An investigation was made to the overall performance of the students as compared to the groups they belong to. The average score between the groups is computed, how far they

differ from their mean and the standard deviation. The average score of the three groups differs significantly due to the anxiety that was hypothesized

Table 2: Performances of the groups

	Normal Start	Easy Start	Difficult start
Mean	7.58	6.48	5.07
Variance	2.90	2.84	2.10
STDev	1.70	1.69	1.45



Figure 2: The Mean, Variance and Standard Deviation of the Groups

Table 2 shows that the group with Normal start performed better with an average score of 7.58 and a standard deviation of 1.70. This is followed by those with easy start with a mean of 6.48, and then finally the difficult start with mean of 5.07 and a standard deviation of 1.45. This confirms our earlier hypothesis that anxiety affects students when difficult

questions are posed first to them. The variance, on the other hand, is with slightly small margins. The performance of the students is not far from the average performance (mostly by < 3 marks). Group A and B differ by about 0.03 and Group A and C differ by about 0.80.

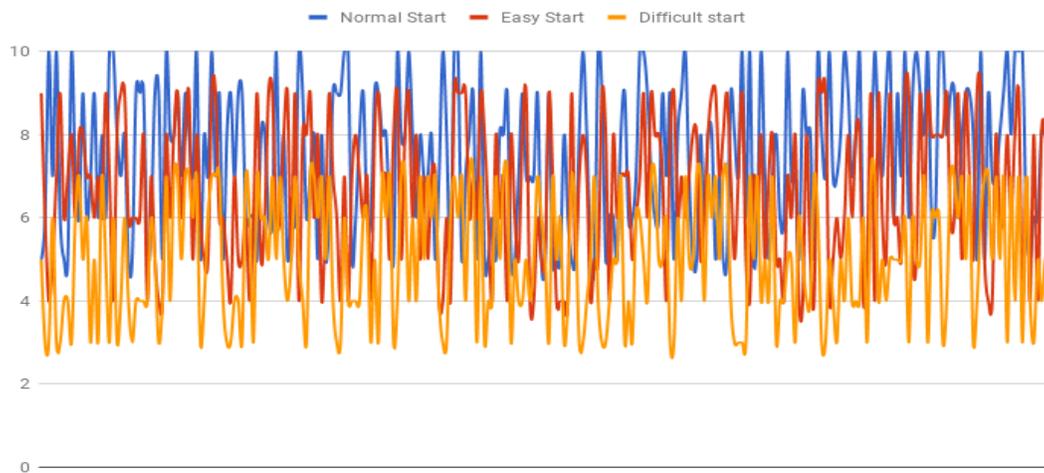


Figure 3: Time Series of the scores of the participants

Figure 3 above shows the distributions of the marks as a time-series graph. It is clear that Group A's performance is better than the other two groups. The spikes show some performances are quite high, sometimes reaching up to the highest score. Group B's performance is average with Group C being the lowest.

CONCLUSIONS

It is clear that randomizing questions has an effect on the performance of students. When the difficult questions were taken first, the performance tends to be the lower of the other methods, and then followed by putting the easy questions first. As an open problem, the study will look at the reasons for this from a theoretical point of view, if there is any psychological theory that backed this up. The study will also design and implement an algorithm that will utilize the difficulty index in generating random questions, so as to minimize the impact on student's performances.

REFERENCES

- Alessi, S. and Trollip, S. (2000). *Multimedia for learning*. 3rd ed. Boston: Pearson.
- Joe, K. and Charles, L.S (1999). "Can We Control Cheating in the Classroom?" *The Journal of Economic Education*, 30(4):331-343, DOI:10.1080/00220489909596090
- Wang, T., Wang, K., Wang, W., Huang, S. and Chen, S. (2004). Web-based Assessment and Test Analyses (WATA) system: development and evaluation. *Journal of Computer Assisted Learning*, 20(1): 59-71.
- Kurt, F., William, S., Gwyneth, B. and Ruth, E. (2004). *Rights and Responsibilities of Test Takers: Guidelines and Expectations*. [online] <https://www.apa.org>. Available at: <https://www.apa.org/science/programs/testing/rights> [Accessed 20 Jul. 2019].
- Marks, A.M. (2008). "Random question sequencing in computer-based testing (CBT) assessments and its effect on individual student performance" (Masters). University of Pretoria
- Pain, D., & Heron, J.L. (2003). WebCT and Online Assessment: The best thing since SOAP? *Educational Technology & Society*, 6: 62-71.
- Ricketts, C. and Wilks, S. (2002). Improving Student Performance Through Computer-based Assessment: Insights from recent research. *Assessment & Evaluation in Higher Education*, 27(5): 475-479.
- Reise, S. and Henson, J. (2003). A Discussion of Modern versus Traditional Psychometrics As Applied to Personality Assessment Scales. *Journal of Personality Assessment*, 81(2): 93-103.
- Steele, J. and Ambady, N. (2006). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, 42(4): 428-436. doi: 10.1016/j.jesp.2005.06.003