

Big Data Analysis of Examinee Multiple Choice Question Examination Log Files

Sami SHABAN

Medical Education Department, College of Medicine and Health Sciences,
United Arab Emirates University, Alain, UAE

Margaret ELZUBEIR

Medical Education Department, College of Medicine and Health Sciences,
United Arab Emirates University, Alain, UAE

1. INTRODUCTION

The Assessment Management System (AMS) at the College of Medicine and Health Sciences (CMHS), United Arab Emirates University has been in use now for five years. 450 assessments of the Multiple Choice Question (MCQ) variety have been conducted over this period. Currently there are over 25,000 unique MCQ questions in the question bank which have been developed by CMHS faculty members for the most part, assessing all levels of the medical curriculum. These questions are stored in the system along with their tags and psychometric data. The aim of this research is to analyze examinee examination behaviors during multiple choice question exams using log file “big data” for the purpose of extracting valid insights leading to evidence-based educational decision-making.

2. METHODS

In addition to storing the questions and examinee answers, the AMS tracks the use of the system by storing every single click performed on the system such as an examinee answering a question or moving to the next question, etc. This is especially useful data collected during examination of examinees. This log file has grown to over 8 million time-stamped entries over the same five-year period. We used this big data to analyze examinee behavior during examinations. The log file is stored in a SQL Server Database System and was partially imported to SPSS for some of the analyses. Around 20 processes were run on the log files to perform the analyses some of which took a few minutes to run on the server and others over 20 hours each (Table 1).

P #	Activity(performed between 1/Dec/2015-31/Dec/2015)	Time to Run Scripts	Number of Records Involved/Created/Remaining
1	Making a copy of the log file	2 minutes	9,021,472 Records
2	Cleaning up the log file from non-student data	10 minutes	8,696,333 Records
3	Cleaning up the log file from duplicate entries	5 hours	7,869,300 Records
4	Creating a table of all relevant exams	3 minutes	490 Exams
5	Creating a table of all Student-Exams with exam time	30 minutes	44,245 Student-Exams
6	Creating a table of all Student-Questions with question time	27 hours	6,780,994 Student-Questions
7	Updating the Student-Questions tables with correct Question/No Using QuestionOrder	20 hours	6,780,994 Student-Questions
8	Updating the Student-Questions tables with correct answers	1 hour	2,030,231 Student-Answers
9	Updating the Student-Questions tables with percent correct and point biserials	30 minutes	2,030,231 Student-Answers
10	Updating the Student-Questions tables with Cognitive Level	7 hours	2,030,231 Student-Answers
11	Creating a table of percent student re-answers	30 minutes	44,245 Student-Exams
12	Updating the Student-Exams table with student correctness	7 hours	44,245 Student-Exams
13	Calculating question length (in characters)	30 minutes	17,276 Questions
14	Creating a table of all unique Student-Answers with question time and length	2 minutes	1,635,220 Student-Answers
15	Creating a table of percent student re-answers for the last 20 percent of answers	45 minutes	44,245 Student-Exams
16	Updating the Student-Exams table with student correctness for the last 20 percent of answers	3 hours	44,245 Student-Exams
17	Updating the Student-Questions tables with re-answer value	2 hours	2,030,231 Student-Answers
18	Creating a table of all Student-Questions for scatter plots	9 hours	6,780,994 Student-Questions
19	Creating a table of all Student-Questions for scatter plots Second Step	8 hours	2,133,336 Student-Answers
20	Creating a table of percent student re-answers by cognitive level	2 hours	170,490 Student-Questions-Cognitive Level
21	Updating the Student-Questions-Cognitive Level table with student correctness for re-answers	6 hours	170,490 Student-Questions-Cognitive Level

Table 1. Data Processes Performed for Analyses

3. RESULTS

450 MCQ exams were analyzed with an average of 99 examinees and 37 questions per exam. The average time spent on exams was 54 minutes with an average of 1.41 minutes per question. The average time spent on questions answered correctly was 71 seconds and on questions answered incorrectly 93 seconds, a significant difference of 22 seconds longer on those answered incorrectly (t-test, p=0.000, n=1,635,220 examinee-answers).

Length of the question in characters was also significantly different between those answered correctly and those answered incorrectly (t-test, p=0.000, n=1,635,220 examinee-answers, a difference of 19 characters out of an average of 329 characters per question).

A typical examinee answering pattern was identified. This involved going through each question in sequence, spending a reasonable amount of time reading each question, answering the question and then reviewing them at the end of the exam while making a few changes to answers (Table 2).

Question No	Question Time(sec)	Question Status	Answer	Correct
1	16	Enter-Answer	1	0
1	7	Answer-Exit		
2	12	Enter-Answer	2	0
2	8	Answer-Exit		
3	8	Enter-Answer	2	0
3	14	Answer-Answer	3	1
3	3	Answer-Exit		
3	4	Enter-Exit		
4	46	Enter-Answer	2	1
4	1	Answer-Exit		
4	1	Enter-Exit		
5	15	Enter-Answer	4	1
5	4	Answer-Exit		
30	7	Enter-Answer	3	1
30	3	Answer-Exit		
1	2	Enter-Exit		
2	4	Enter-Exit		
3	33	Enter-Exit		
4	66	Enter-Answer	3	0
4	5	Answer-Exit		
4	8	Enter-Exit		
5	17	Enter-Exit		
6	16	Enter-Exit		
7	9	Enter-Exit		
8	2	Enter-Exit		
9	8	Enter-Exit		
10	11	Enter-Exit		
11	25	Enter-Exit		
11	3	Enter-Exit		
12	27	Enter-Exit		
13	8	Enter-Exit		

Table 2. Identified typical examinee answering pattern

A significant but very weak correlation was found between average time spent on question and length of question in characters ($r=0.280$, $p=0.000$, $n=1,635,220$ examinee-answers) (Figure 1).

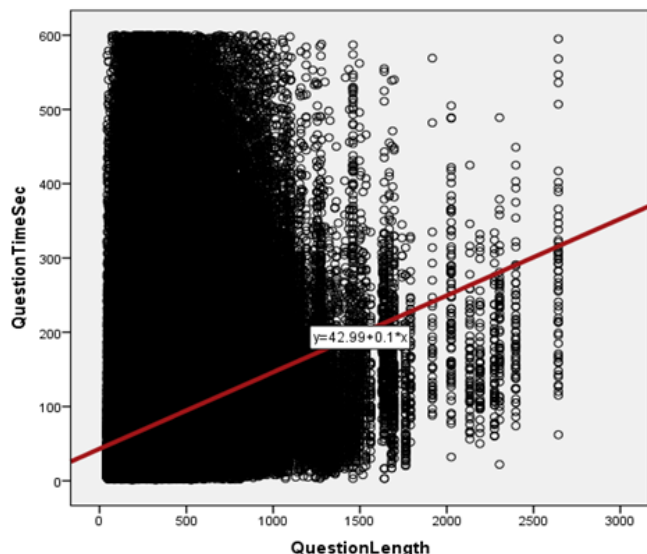


Figure 1. Linear regression of time spent on question and length of question in characters

We found significant differences with regard to the Bloom's Taxonomy classification of questions (cognitive level from 1: Knowledge to 6: Evaluation) in correctness, point biserial, question length and time spent on questions. As cognitive level increased, correctness decreased, point biserial significantly decreased, question length significantly increased, and time spent on question significantly increased (Table 3).

Cognitive Level	Number of Questions	Average Mark %	Average Point Biserial	Average Question Length in Characters	Average Question Time in Seconds
Unclassified	688,628	67.1	0.360	307.4	79.3
1 Knowledge	478,136	72.0	0.331	270.1	65.3
2 Comprehension	255,454	68.0	0.316	374.3	81.9
3 Application	96,393	63.0	0.282	426.5	90.3
4 Analysis	65,663	72.4	0.273	514.6	90.8
5 Synthesis	38,868	72.2	0.287	530.0	87.3
6 Evaluation	12,078	68.3	0.299	483.5	81.9
Total/Average	1,635,220	71.8	0.352	328.9	76.9

Table 3. Comparisons regarding Bloom's Taxonomy classification of questions

Finally, with regard to re-answering questions, 13.5% of examinee-questions were re-answered: 5.4% from incorrect to correct and 2.5% from correct to incorrect. This is very similar to re-answering percentages found in another study [1].

Number of Exams	Average Number of Questions	Average of Incorrect To Correct Change	Average of Correct To Incorrect Change	Average of Correct To Correct Change	Average of Incorrect To Incorrect Change
450	37.01	2 (5.41%)	0.92 (2.49%)	0.49 (1.33%)	1.6 (4.32%)
Last 20% of Qs	7.64	0.92 (12.02%)	0.49 (6.38%)	0.26 (3.4%)	0.86 (11.29%)

Number of Exams	Average of Re-Answered	Average of Re-Answer Once	Average of Re-Answer More	Average of Correct Re-	Average of Correct Re-
450	5.01 (13.54%)	3.34 (66.72%)	1.67 (33.28%)	3.84 (76.6%)	1.37 (27.43%)
Last 20% of Qs	1.16 (15.23%)	0.8 (68.4%)	0.37 (31.6%)	0.84 (72.11%)	0.29 (25.06%)

Table 4. Percent of question re-answering

4. DISCUSSION

We found only one article discussing analysis of examination log files [1]. They have conducted much fewer examinations than we have and so their sample size is much smaller. Nevertheless, they discuss the issue of residents changing their responses on 12% of questions and that changes from incorrect to correct was almost twice as likely as changes from correct to incorrect. These are very similar results to ours with regard to examinees changing responses (Table 4).

5. CONCLUSION

Analyses of large log files can give insight into the behavior of examinees during online exams. For example, examinees spend more time on questions they answer incorrectly and more time on higher cognitive level questions. Reviewing and re-answering questions has a small benefit to the examinee's mark. In addition, dealing with big data involves dealing with several systems that can handle large data storage and analysis, and hours of processing time using these systems yielding many analyses that can be performed. Analyses done with big data are always highly significant if a difference exists because of large power of the large sample size and therefore correlations may be spurious.

6. REFERENCES

[1] Hess BJ, Lipner RS, Thompson V, Holmboe ES, Graber ML. **Blink or Think: Can Further Reflection Improve Initial Diagnostic Impressions?** Academic Medicine. 2015; 90(1): 112-118.