How question shuffling influences correlation of student
text item selections in computer-based tests

Bee-Sung Kam

Abstract

Handling Computer Based Test (CBT) in large classroom requires more of test reliability. Because of limited space assigned to each student, computer monitors are positioned close together so that reliability of exam goes under question also the environment makes student vulnerable to copy from their neighbors. A method to deter students from copying answers of their colleagues is to shuffle items in multiple choice questions. Objective of this study is compare student selected items and questions before and after randomizing assessment of multiple choice test to emphasize that reordering questions can set an environment to discourage students from copying answers of their colleagues in an exam room. This study used existing data from students enrolled in circulation system assessment of Pusan National University School of Medicine over two academic semester year of 2013 and 2016 and measured for correlation coefficient of pre and post randomization. Statistically significant relationship between selected items were found in favor of item shuffling in CBT.

Keywords: Randomizing, Computer Based Test, Cheating Deterrence, Computer Based Assessment

1. Introduction

Fact that student cheat to attain opportunity is a well known method of adoption [1]. “How student cheat” is a question that is largely neglected [2]. We define cheating in this paper as “copying from another”. In medical school cheating behaviors including copy from others have been detected [3]. Ability to copy answers from another is not necessarily because of student’s frustration in their exam. A normal settings may force student to be vulnerable to copy from neighbors or being copied themselves. Reduction method of cheating have been studied and implemented using seating rearrangements, or displace friends to reduce drawing
attentions which are very common and useful method [4]. In an exam room relocating good students in order to cast a vigilant eyes also is essential. Computer methods of deterrence for these kinds of behaviors are to randomly positioning students on their undetermined seats so that environment itself is a method of deterrence [5].

Large scale classroom examination is often limited by space provided for each examinee so that it is not difficult for students to see answers of other students who have sited in front, diagonal-front and sides. Also because of limited vision, distance and movements of head and body during assessment we observed that all students can copy only from five seating positions mainly (L: beside in left, LF: diagonal left-front with distance, F: front, RF: diagonal right-front with distance, R: beside in right) and only if possible copying answers is a matter of wish by trying to look in different directions and only mentioned positions. Privacy screen filters have been used to control vision [6] but only 54% of student reported thinking that filters were effective. In other hand filters tend to deem color of images and because vast majority of images in medical schools are in deep color, deeming color for example in a skin tissue would confuse student for a distinguished answers therefor monitors without filters are more appreciated.

Computer based test software is being introduced for the ability to randomize generated questions for student examination. Although scrambling is a challenging issue for data management because unscrambling test items must be dealt with in scoring process.

This have been achieved particularly in tests that question items are not large enough to be replicated for randomize testing. This paper presents a study to achieve how question shuffling influences correlation of student selected items in computer-based tests. Data from multiple-choice questions compared in pre and post randomization of CBT and resulted in reduction of item selection correlations.

2. Method

This study used existing data from CBT assessment of second grade students enrolled in assessment of circulatory system course in Pusan National University School of Medicine over two academic semester year of 2013 and 2016. Data collected from participant included test score, multiple question selection number, question difficulty level, test grade and seat number. Level of difficulty for compared assessment data were similar in pre and post assessment measured in 73% and 67% respectively. Table.1. illustrates number of participation and Fig.1. illustrates design of the study.
Data of total 248 student were selected and distributed to five categorical sitting positions including L: beside in left, LF: diagonal left-front with distance, F: front, RF: diagonal right-front with distance, R: beside in right. During examination a student could possibly view only five screen monitors of neighboring colleague mainly (L, LF, F, RF, and R). Participants of both data sets were not aware of any future data analysis for cheating deterrence or measuring of copying answers from neighbors. There were no any punishment for turning eyes in order to view colleague monitors. Data of students from front and ending row were eliminated firstly because of unavailability to view of any screens, and secondly because of existence of semester intervals and unequality in data selection.

Data participation rate for pre and post randomization were 88% and 90% respectively. Students took test in a classroom furnished with 70 desks each facilitated with two desktops and 21 inch wide monitors. Two students randomly positioned in each desk with a distance of 50cm to the left and right neighboring monitor, almost 90 cm to monitors at front and 120cm to diagonally-left or diagonally-right monitors. A guideline has been introduced to student according to [7]. (1) smuggler: usage of forbidden material; no material and devices allowed into exam room (e. g. USB, Phone, etc-materials gathered before entrance), (2) The tourist: randomizing question item and items within question, also student seats are selected randomly (3) the incontinent: restroom breaks: one can go to the restroom only if he/she submits his/her exam (4) The impersonator: one with false id; students use unique recognizable id, (5) The hacker: accessing assessment system: assessment provided with an offline system authorized only to assessment assistant. (6) the storyteller: cheating by asking student whom have taken exams already (e.g. clinical skills assessment), speaking in exam room was prohibited, (7) The air traffic controller: assess to item discussions: newly developed items created yearly for each exam and question opened to discussion after each examination, (8) the collaborator: groups who providing unauthorized assessment materials and (9) The empathizer: previous school year failed students:

<table>
<thead>
<tr>
<th></th>
<th>Pre-Randomization</th>
<th>Post-Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>1st semester year of 2013</td>
<td>1st semester year of 2016</td>
</tr>
<tr>
<td>Grade</td>
<td>2nd grade</td>
<td>2nd grade</td>
</tr>
<tr>
<td>Lecture</td>
<td>Circulatory system lecture assessment</td>
<td>Circulatory system lecture assessment</td>
</tr>
<tr>
<td>N</td>
<td>128</td>
<td>120</td>
</tr>
<tr>
<td>Level of difficulty</td>
<td>73.99</td>
<td>67.33</td>
</tr>
</tbody>
</table>
previous school year examination were open to public and discussed in open class. also new item banks were made yearly. (10) Robin Hood: well-intentioned instructors: question were responded after exam, also assistant without knowledge of examination selected to be assisting examination.

Pre and post CBT assessments executed using developed software of Pusan Medical Assessment System (PMAS) program implemented by medical education unit development group, ran in windows environment and viewed in one page style multiple questions granted with sliding bars, check buttons for later view, answer buttons, back-front functionality to and from multiple question/short answers, drawing pen and note.

N=109 data set from each group of pre and post assessment selected and compared to five distinguished seating position (L, LF, F, RF and R). Correlations of 109 seats with five seating positions measured statistically according to selected item of multiple choice for each type of neighbors. The average results were then calculated and compared using Pearson’s correlation of each type of neighbors.

Statistical significance was set at p=0.05. We used Cohen’s to estimate the effect size and therefore practical significance of statistically significant differences, adopting the convention where values of >0.1 , >0.05 and <0.05 indicate less-cheating, possible-cheating and cheating effects respectively.
3. Results

Assessment data were examined before and after question randomization with respect to year in school and sitting positions. Sitting position categories included, L: beside in left, LF: diagonal left-front with distance, F: front, RF: diagonal right-front with distance, R: beside in right. Table one illustrates comparison of pre and post data sets with respect to seating positions. Table 1. illustrates pre and post randomization of CBT assessment.

[Table 2] Illustration of pre and post shuffling in CBT with respect to seating position (N=109) p=0.001

<table>
<thead>
<tr>
<th>Type</th>
<th>L</th>
<th>LF</th>
<th>F</th>
<th>RF</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-randomization</td>
<td>.781</td>
<td>.785</td>
<td>.782</td>
<td>.785</td>
<td>.781</td>
</tr>
<tr>
<td>post-randomization</td>
<td>.777</td>
<td>.784</td>
<td>.774</td>
<td>.775</td>
<td>.776</td>
</tr>
</tbody>
</table>

In position LF value of correlation coefficient before randomizing assessment were 0.785 while in post randomization coefficient value reduced to 0.775, differences of 0.010 in coefficient value.

In position F value of correlation coefficient before randomizing assessment were 0.782 while in post randomization coefficient value reduced to 0.774, differences of 0.008 in coefficient value.

In position R value of correlation coefficient before randomizing assessment were 0.781 while in post randomization coefficient value reduced to 0.776, differences of 0.005 in coefficient value.

In position L value of correlation coefficient before randomizing assessment were 0.781 while in post randomization coefficient value reduced to 0.777, differences of 0.004 in coefficient value.

Result Fig.2. illustrates reduction of coefficient value for each seating position in post-randomization of assessment which means that copy answers from neighboring students have been reduced in post-randomizing assessment.
How question shuffling influences correlation of student item selections in computer-based tests

4. Discussions

A method to deter students of copying answers from colleagues is to randomize questions and shuffle items in multiple choice questions. Although it is unclear how much cheating actually happens in exam room, possibility of looking at colleague’s exam screen produces rumor that somehow somebody is being able to copy one another’s answer which gain unfair score and open thrust of the heat of compatibility [8]. This study gives a solution in which enforces the idea of “copying is difficult” which psychologically make a more relaxing environment in exam room and student feel more safe. In reality questions are in random position so that firstly it is difficult to position a question within questions, secondly question items are shuffled so that locating an items and finding answer is not easy, to do so one must gaze into another screen with open eye-bowl to find out which written item is matches answer, thirdly because checking answers appears only in numbers the cheater must guess the number because in more than 50cm distance a small check point is almost invisible.
A study on nursing examination [9] reveals that 10% of copying answers from colleague by looking at the work of another. While in reality looking at another screen is hard to justify for dishonesty we introduce post randomizing assessment which is a method that ensures looking to be free, because by only looking it is difficult to get to know some answers to a specific question on a moving screen. The “looking” must be more of a gaze with both eyes open toward a screen which can not be easy in a class that has supervisor looking.

Another method of cheating deterrence is by making guidelines a 5 values promoting academic integrity, honesty and trust, fairness and respect. Firstly student engagement would be an important factor in promoting ‘buy-in’ for the recommendations and secondly addressing suspected cheating behaviors as a recommendation would enforce demand for deterrence [9]. Limitation of such a method is student won’t participate in such a development also it is difficult for an individual to be responsive in a class for reporting dishonesty before graduation. Study of Fendler [10] represents a higher percentage of probable copying when student assign their own seat. They concluded that separating test-takers from their acquaintances can reduce copying effectively in a multiple choice exam. In this study we randomly assigned student to their seats and it may slightly effect overall result of this study. We conclude that environment setting in important in classroom examinations. Randomizing questions and shuffling question items are a valid method to perform fixed question examination is a large classroom. Reliability and fairness of CBT can vary depend on environmental settings and not necessarily to student frustration in earning score as a method of deterrence.

References


