**Abstract:** The purpose of this study is to compare CBT (computer-based testing) and paper testing. In an area like regular expression, if tested by CBT, the students can simulate and check their answers during the examination time. Therefore, the authors of this study hypothesized that in CBT, compared with paper-testing, 1) the students will get higher scores, 2) they will anticipate their scores more accurately, 3) the rate of blank answer will rise, and 4) the students will have better impression of the examination. In order to prove this hypothesis, the authors of this study conducted a one-hour class of regular expression to 40 students who were from different universities and in different school years. After the class, the students were divided into two 20-member groups and took the same 12-question examination, one group by CBT and the other by paper-testing. The students also had to answer two questionnaires, one before and the other after the examination, which inquired the students’ personal attributes and their confidence levels in their answers. The result showed that the CBT group (5.85 points out of 12) had slightly higher scores than the paper-testing group (5.05 points out of 12). This difference was of no statistical significance. However, in the rate of blank answers, the CBT group (4.58 points out of 12) showed a significantly higher rate than the paper-tested group. This seems to suggest that, in CBT, students “work with each question until they solve it, and answer only when they know it is correct.” At the same time, because they took too much time working with the questions in the beginning of the examination sheet, they lost time for the rest of the questions. As a result, the tendency was that there was no significant score difference between the CBT group and the paper-tested group.
1. Introduction

1.1. Background

Computer-Based Testing (CBT) has many strong points such as: 1) the students can choose the examination time and place, 2) scoring is easy, 3) the students get immediate feedbacks as to the correctness of their answers, and 4) the data of the test results is easily available for various purposes.

In addition, in CBT, it is possible to simulate one’s answer on the computer. For example, Muraki & Hagiwara (2002) stated that CBT “is very suitable for language-skill exams, computer-literacy exam and other performance-based exams” because it can provide problems utilizing multi-media functions. Aoki (2001) developed a system for C-language exam in which the student answers on the Web and gets scored immediately.

1.2. Purpose of this study

The purpose of this study was to compare CBT and paper-testing. The working hypotheses were as follows.

[Hypothesis 1] The students will get remarkably higher scores in CBT, in which they can simulate their answers on the computer screen, than in paper-testing.

[Hypothesis 2] In CBT, the students will be able to anticipate their scores more accurately than in paper-testing.

[Hypothesis 3] In CBT, there will be higher rate of blank answers.

[Hypothesis 4] The questionnaires will show that the students had more favorable impression on CBT than on paper-testing because the former allows them to learn the subject matter during the exam.

2. Method

40 university students (20 male, 20 female) who had chosen mathematics for their entrance exams of their universities (hence, have basic mathematical skills) were asked to take a class of regular expression. The group Composition is shown in Table 1.
Table 1: Subject Group Composition

<table>
<thead>
<tr>
<th>Gender</th>
<th>University</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Public</td>
<td>Education</td>
</tr>
<tr>
<td>Female</td>
<td>Private</td>
<td>Economics</td>
</tr>
<tr>
<td>1st-year</td>
<td>9 schools 16 students</td>
<td>Law</td>
</tr>
<tr>
<td>2nd-year</td>
<td></td>
<td>Healthcare</td>
</tr>
<tr>
<td>3rd-year</td>
<td></td>
<td>Engineering</td>
</tr>
<tr>
<td>4th-year</td>
<td></td>
<td>Life Science</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Literature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Natural Science</td>
</tr>
</tbody>
</table>

2.1. Exam of regular expression

All the 40 students were assembled in one room and took a 60-minute class on regular expression, which was conducted on March 8th, 2010, from 3:00pm to 4:00pm.

In the class, one note-type PC was given to each student. This PC had a software program named “Regular Expression Quiz” with which the student could solve a problem of regular expression by simulating the answer. “Regular Expression Quiz” will be explained in Chapter 3 (Developed System).

2.2. Pre-exam Questionnaire

Immediately after the class, in the same room, the pre-exam questionnaire was given to the students. In the questions (10 total), except for Questions Q3 and Q4, the students were asked to select one from 5 answers ranging from “Yes, I feel that way” to “No, I don’t feel that way.”

Q1. I have learned about regular expression before.
Q2. I have experienced making a computer program before.
Q3. Please circle the OS programs (plural circles OK) you have ever used.
   Windows  Mac  Linux  Unix  Others ( )
Q4. Please circle the software programs you have ever used.
   Word-Processing  Table-Calculation  Presentation  Database
Q5. The class was easy.
Q6. The class felt short.
Q7. Regular expression was easy.
Q8. Regular expression was interesting.
Q9. Exercises in “Regular Expression Quiz” were interesting.
Q10. The character figures by AA that appeared in the software “Explanation of Level XX” were interesting.
2.3. Break

The students were given a 10-minute break after the class, during which Group B (paper-tested) was moved to an adjacent conference room.

2.4. Test of Regular Expression.

An exam was given to test the knowledge from the class and its easy application. There were 12 questions. Since they were printed on double-side A-4 papers, each student was given 6 question/answer sheets.

In this article, the pre-exam questionnaire for Group A is called a1, this exam itself (for Group A) A, the post-exam questionnaire a2. For Group B (on paper-testing) the equivalents are called b1, B and b2.

We made the following hypotheses.

[Hypothesis 1] Group A’s scores will be higher than Group B’s.
[Hypothesis 2] Group A students will anticipate their scores more accurately than Group B.
[Hypothesis 3] Group A students will have higher rate of blank answers. (This is because in CBT, people tend to answer only when they know it is correct by simulating it on the computer screen.)
[Hypothesis 4] When the results of the two (pre/post-exam) questionnaires are compared in each group, Group A (a2-a1) will show a bigger improvement in their impression on the exam than Group B (b2-b1).

While both Group A students and Group B students wrote their answers on the answer sheets, only the former were allowed to open and use the software program called “Regular Expression Quiz.” CBT in this experiment only means that the students (in Group A) were allowed use it during the exam in order to work with the problems; they had to write their answers on the answer sheets. The reason for this setup was to eliminate the possible variance of scores due to differences in manual processes between writing answers on the answer sheet and entering them on the computer screen (which is technically possible in CBT).

2.5. Post-exam Questionnaire

After the test, in the same room, the post-exam questionnaire was conducted. There were ten questions. Except for Question 10, the students were asked to answer by selecting one from 5 optional answers ranging from “1. No, I don’t feel that way” to “5. Yes, I feel that way”.

Q1. (Please answer only those who used the software program Regular Expression Quiz during the exam.) It was good that I could use a PC during the exam.

Q2. (Please answer only those who were not allowed to use Regular Expression Quiz.) It would have been better if I could use a PC during the exam.

Q3. I learned the subject not only during the class but also during the exam.
Q4. While taking the exam, I could remember what I had learned during the class.
Q5. The exam questions were easy.
Q6. I felt that the exam time was short.
Q7. I felt that regular expression was easy.
Q8. I felt that regular expression was interesting.
Q9. I felt that the exam questions were interesting.

3. Program Developed

3.1. System used during the class

During the above class and the exam for the CBT group, a software program, with which the user can simulate and test regular expressions, was used. The URL of the website is http://aastory.info/reg/index.html. The screen of the software is shown as Fig. 1.

3.2. The system used for the exam

Class A (the CBT group) took the exam using basically the same system, except that the functions of showing “Hint” and correct result sentence are eliminated.

4. Exam Result

Table 2 shows the students’ exam scores in top-down order, contrasted with the scores anticipated by the students themselves right after the exam.
4.1. Scores

There was no significant difference between Group A (who could use PC during the exam/ average 5.85 points) and Group B (who could not use PC/ average 5.10 points). Hence, in this experiment, Hypothesis 1 was not supported. (t-test, P (T<=t), 0.202 unilateral). From this result, the authors of this study assumed that the CBT-group students spent too much time on each question and lost time for the latter questions. The blank answer rate is shown in the next chapter. The actual/ anticipated score differences were also not significant between Group A and Group B. (t-test, P (T<=t), 0.270 unilateral). Hence, Hypothesis 2 was not supported either.

4.2. Blank Answer Rate

As to the blank answer rate, there was significant difference. The same exam of 12 questions being given to both groups (with 20 students each), there were 240 answer boxes in total. Group A, in total, left 55 answer boxes blank, whereas Group B left only one blank answer. (t-test, P(T<=t), 0.0014 unilateral). The two groups’ blank answer rate differences are shown in Figure 2.

Fig. 2: Two groups’ blank answer rate differences
The CBT group shows significantly higher blank-answer rate. In addition, it goes up significantly toward the latter questions in the exam. This result supports Hypothesis 3.

4.3. Observation

Between the CBT group and the paper-testing group, while no significant difference was seen between their average scores, there was remarkable difference between their blank answer rates. This result supports the following assumption; in CBT-style exam of regular expression, while there is no visible improvement from the paper-testing group as to their total scores, the students showed more “guts” in solving the problems and answered only when they had correct answers.

4.4. Questionnaire Result

Hypothesis 4 is now discussed. The two questionnaires (Pre-Exam and Post-Exam) had two questions in common: “7. Regular expression was easy” and “8. Regular expression was interesting.” The Pre-Post changes in both groups’ answers to these two questions are shown in Fig. 3 and Fig. 4.

![Fig. 3 “Regular expression was easy”](image)

In Fig. 3, the CBT group found regular expression easier after the exam than before the exam. This suggests that they could learn regular expression by simulating answers during the exam, which made it look relatively easier to them. As to Fig. 4, the paper-tested group shows more improvement in their impression, “Regular expression is interesting.” The paper-tested group may have felt that regular expression was difficult but interesting.

5. Conclusion

In this study, the following hypotheses were posited.

[Hypothesis 1] Group A’s scores will be higher than Group B’s.
[Hypothesis 2] Group A students will anticipate their scores more accurately than Group B students do theirs.

[Hypothesis 3] Group A students will have higher rate of blank answers. (This is because in CBT, people tend to answer only when they know it is correct by simulating it on the computer screen.)

[Hypothesis 4] When the results of the two questionnaires are compared in each group, Group A (a2-a1) will show a bigger improvement in their impression on the exam than Group B (b2-b1).

In order to prove these hypotheses, 40 students were asked to take a one-hour class on regular expression. After this, they were immediately divided into two 20-member groups. They took the same exam of regular expression, one by CBT and the other by paper-testing. They also had to answer the same two questionnaires, one before and the other after the exam. The result did not support Hypothesis 1 and 2. Hypothesis 3 was proved to be correct. As to Hypothesis 4, the result seemed to support it, at least partially.

References