Strict Order vs. Flexible Order Navigation in Online Testing

Ekaterina Vasilyeva, Paul De Bra, Mykola Pechenizkiy

Database and Hypermedia Group, Department of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands

e.vasilyeva@tue.nl, debra@win.tue.nl, m.pechenizkiy@tue.nl

Abstract. This paper presents the results of experimental studies organized during recent midterm examinations in the form of online assessment with multiple-choice questions. Three kinds of tests were organized. In the first test the students had to answer the questions in a strict order. In the second test students had a possibility to answer the questions in a flexible order, to revisit earlier answers and revise them as well. In the third test students answered to the questions in a strict order, but received additional information about time recommended for answering each question. Our results suggest that flexible order navigation helps the students to evaluate the relative difficulty of the questions and to better distribute the limited time they have between these questions. However, the results also suggest that providing the time-related recommendations in a test with a strict order also allows the students to estimate and distribute the time needed for each questions correctly. This allows teachers to take advantage of the strict order, for instance to give immediate elaborate feedback that may help students in answering questions that appear later in the same exam.

Keywords: online assessment, navigation, e-learning.

1 Introduction

An increasing popularity of different e-Learning environments suggests to many educators to reconsider the choice of student evaluation procedures. Online testing organized in a form of multiple-choice test becomes more and more common and is often favored over written and face-to-face oral examinations. It is now used not only in e-Learning, but within blended learning for self-evaluation of the students and for real assessment, i.e. partial or final exams. There are clear advantages of the online assessment such as the full or partial automation of the assessment process, providing students with immediate (or more quick) and elaborated (possibly adaptive) feedback, easier re-use of the assessment tasks and questions, collection of the data about the student performance (that can help in more detailed analysis of the students’ misconceptions, problems, etc), storing of the students works, etc. However, online assessment of the students still has problematic aspects that have to be addressed; those e.g. include usability, security, navigation and other issues. In this paper we
study the problem of navigation during online testing used for real examination of the students (mid-term or final exams).

In computerized adaptive testing the selection of the next question is usually based on the student's answer to the previous question(s) [5]. This assumes a strict (but personalized) order of taking the questions. A strict (but not necessarily personalized) order is also commonly used in tests where immediate feedback is provided after each question, especially when that feedback may be helpful in answering a later question [4]. Our recent studies have shown the advantages of immediate (personalized) feedback presentation in online assessment [3]. However it is not immediately clear whether an implied strict order of navigation has (dis)advantages for the students. In our practice we often get requests from students for having a flexible navigation through the tests. Students have been used to flexible navigation in plain written exams (where they can scan the entire exam first, skip some presumably more difficult questions and come back to them later, and put last minute changes if they like).

In this paper we analyze and compare two navigational strategies in online assessment: strict-order navigation, when the student answers the questions one after another and has no possibility to return to the previous questions and revise his/her answers, and flexible-order navigation throughout the test, when the student can revisit his earlier answers and revise them.

We present the results of the experimental study organized during midterm exams in the form of online assessment with multiple-choice questions. In the first test, students had to answer the questions in a strict order without any possibility to return to the previous questions. In the second test students had a possibility to answer the questions in a flexible order, to revisit earlier answers and revise them as well. In the third test the students answered to the questions in a strict order, but the information about the time recommended for answering each question was provided.

In this paper we analyze the students' navigational paths and strategies for passing the tests, the individual and average scores and time they spent, the number of the cases when the students changed their answer in the tests with flexible-order navigation and how the students followed time-related recommendation while answering to the test with strict order navigation.

The data obtained from the experiments was also used to demonstrate the potentials of using process mining [6] techniques for educational data analysis, which we discussed in [7].

The rest of the paper is organized as follows. In Section 2 we describe the experimental procedure. In Section 3 we present and discuss the results of our experiments. We conclude our paper with a brief discussion and directions of further research in Section 4.

2 Method

We studied navigational strategies in online assessment within the series of the experiments organized as mid-term exams at the Eindhoven University of Technology in the form of on-line assessment with multiple-choice questions. The tests were
organized as two partial exams for Databases (DB) and one partial exam for Human-Computer Interaction (HCI) (bachelor-level courses). Each test consisted of multiple-choice questions on the topics of the course (DB tests – 10 questions, HCI test – 15 questions).

During the design of the questions for the quizzes the teachers tried to invent incorrect but believable answers and were taking into account typical mistakes students make and misconceptions they may have. The questions were aimed at assessing the knowledge of the concepts and the development of the necessary skills (like understanding of the basic usability rules and problems such as consistency, mapping (between interface and real world), response time problem in HCI test or like computing a canonical cover or translating between plain English and SQL in the DB tests).

The Sakai learning system, and particularly a slightly modified Mneme testing component was used to organize the online testing. The students took the tests individually for a partial grade for the course (15% of the final grade for each midterm examination).

The tests were taken simultaneously by all students of the course in the same place. The students used their own laptops and were allowed to use any kind of possible sources of information except for communication tools. The teacher with 2 assistants was present in the class to observe and to assist the students.

Short instructions about the number of questions in the test, the grading scheme, the functionality of the system (including the possibilities of navigation) were given to the students before each test. The students started each test simultaneously; time was limited to 80 minutes. For each question of the test students had to give their confidence (certainty) level (high vs. low) for the selected answer. Our previous studies demonstrated that knowledge of the response certitude (specifying the student’s certainty or confidence of the correctness of the answer) together with response correctness helps in understanding the learning behavior and allows for determining what kind of feedback is more preferable and more effective for the students thus facilitating personalization in assessment [15].

In the first DB test (DB1) students answered to the questions in a strict order. In the second DB test (DB2) students had a possibility to answer the questions in a flexible order, to revisit (and revise if necessary) the earlier answers. In both DB1 and DB2 tests the information of the time remaining until the end of the test was provided throughout the text. In the HCI test the order of the questions was strict, but besides the remaining time indication additional time-related information was given. The students received information about the time recommended for answering the question; and also information about the expected time needed to answer to the remaining questions of the test.

86 students participated in the DB1 test, 95 students in the DB2 test, and 73 students in the HCI test.

For every student and for each question in the test we collected all the possible information, including correctness, certitude, grade (determined by correctness and certitude), time spent for answering the question, and for the DB1 test whether an answer was checked for correctness or not, whether detailed explanation was requested on not, and how much time was spent reading it, and for the HCI test
whether a question was skipped, revisited, whether answer was revised or the certitude changed.

3 Results

In Fig. 1 and Fig. 2 we present the averages across all the students for the time used for answering each of the 10 questions with the strict linear order (left) and flexible (right) navigation (Fig 1) and the corresponding average error rate for each question (Fig 2). Dashed lines correspond to the average time spent and relative difficulty of the questions (with respect to average error) while seeing each question for the first time.

It can be clearly seen from the figures that the students estimated the difficulty of the questions more correctly in the DB2 test with flexible order navigation (right). In the strict order tests the students spent more time for the questions which were answered correctly most of the time (e.g. for questions 7 and 8 students spent more than 400 seconds on average, while the average error for these questions was lower than 0.2), while for more difficult questions students spent less time (e.g. for questions 2 and 9).

Fig. 1. Average time used for answering each of the 10 questions with the strict linear order (left) and flexible (right) navigation. Dashed line corresponds to the average time spend while seeing each question for the first time.

Fig. 2. Average error (corresponds to the difficulty of a question) for each of the 10 questions with the strict linear order (left) and flexible (right) navigation. Dashed line corresponds to the “temporal” average error for each question for the first time answer (if provided, otherwise not included in computing an average).
In the DB2 test where students had a possibility to answer the questions in a flexible order the students spent more time for the most difficult questions (e.g. for questions 4, 6, and 8 they spent more than 600 sec on average, while average error was higher than 0.4).

Figure 3 illustrates the output of the dot chart analysis of the flexible-order online assessment. All the instances (one per student) are sorted by the duration of certain actions during online assessment (reading and answering the question and navigation to the list of questions). At the figure on the left points in the ochre and greed/red color denote the start and the end (passed/failed) of the test. Triangles denote the moment when the student submits an answer or just navigates to another question. Green triangles denotes correct responses with low (low certainty correct response (LCCR) – light green) and high (high certainty correct response (HCCR) – dark green) certainty, red triangles correspondingly – wrong responses (light red – low certainty wrong response (LCWR), dark red – high certainty wrong response (HCWR)), white triangles – the cases when the student navigates to the next question without providing any response. The blue squares show the moments when the students navigated from the list of the questions (menu) to a question of the quiz (or just submitted the whole test).

Fig. 3. DCA of test with flexible order navigation.

We can clearly see from the figure that most of the students answered the questions one by one, and provided more correct answers for the first questions of the test than for the last questions. They used the possibility to flexibly navigate mainly at the end of the test: students navigating to the list of the questions and then to the different questions from the list. It can be also clearly seen that only few students read and
skipped some questions, not providing their answers first, and then returning to those questions back to provide an answer.

In Figure 4 we can see the situations when students revisited the questions. Points in yellow correspond to the situations when correctness of the answers did not change, and points in red and green correspond accordingly to changes to wrong and correct answers. We can see that in a very few cases the correctness was changed and when a change occurred they either did not have any effect (when wrong answer was changed to another wrong answer) or the opposite changes from right to wrong or from wrong to write had similar frequencies, thus not significantly changing the end results..

In total, the students performed 673 revisits\textsuperscript{1}, among which 376 times it was just browsing without changing an answer. Among 296 revisits after which an answer change took place, the number of correct changes (skipped answer to correct or wrong to correct) was 111 and the number of unsuccessful changes (skipped answer to correct or wrong to another wrong) was 173 plus from correct to wrong – 12. Thus, the more difficult questions were reconsidered more often and in many cases it was done successfully. The dashed line in Fig. 2 suggests that there is no clear pattern as for who answers the difficult questions from the first attempt – stronger or weaker students. The further detailed analysis is needed to uncover such dependencies if any.

\textsuperscript{1} A number of deviations from the strict linear order increased linearly with every next question. The number of revisits varied from 0 to 28 per student (80\% of which - from 1 to 18).
In these tests students were asked to indicate their confidence or certainty of their answer. Besides the answer correctness we analyzed how this certainty of answers was reconsidered. For both situations when the students changed certainty from low to high (30 cases) and when students changed certainty from high to low (15 situations) the relation between the number of correct and incorrect responses was the same (63-65% correct vs. 35-37% incorrect). This shows that re-considering the certainty does not necessarily increase the accuracy of the student’s prediction of the correctness of the answer.

Fig. 5 and Fig. 6 give us a clear demonstration of how additional information about the time recommended for answering the current question at consideration helps the student to distribute the time between the questions. Although, the students in reality spent less time on average (lower curve in Fig. 5) in absolute numbers than was recommended for answering the question (upper curve in Fig. 5), the results clearly show that the students followed our recommendations and spent more time for more difficult questions and less time for easier questions.

**Fig. 5.** Average time used (bottom line) vs. recommended (top line) for answering each of the 15 questions

**Fig. 6.** Average error for each of the 15 questions (corresponds to the difficulty of a question; the higher the value the more difficult the question is).
The average error numbers (Fig 6) also clearly depict that the students made more mistakes for the most difficult questions and less mistakes for the easier ones. Average error rate was unexpectedly high for most of the difficult questions. This can be also partially explained by the fact that students spent on average less time than was recommended and seriously underestimated the difficulty of some questions or simply gave up and made a guess (see Fig 5).

4 Discussion and Conclusions

Navigation of the students in e-Learning systems has been actively studied in the recent years. One of the directions in this research is social navigation that is based on the analysis of the community navigational strategies and denoting/recommendation of the most visited content for students’ navigation support [8].

In general, the researchers try to discover individual navigational styles of the students in order to reduce cognitive load and losses of the students, to improve usability and learning efficiency of e-Learning systems and support personalization of navigation [1][2].

For example, in [2] navigation of the students in three learning environments with the same learning materials, but with three different navigational possibilities and structure were compared: linear, star and interconnected.

Adaptive navigation support in e-learning demonstrated its advantages for improvement of the learning outcomes, increasing the speed of learning, encouragement of non-sequential navigation, as well as increasing in the amount of students’ work with non-mandatory educational content [9].

In the context of personalized adaptive assessment it is not immediately clear whether an implied strict order of navigation results in any advantage or disadvantage for the students.

In pedagogical literature there is a lot of discussions about the design of multiple-choice questions [10, 11], different forms of assessment in e-learning [12, 13] and its importance, however there is a dearth of pedagogical guidelines for practitioners on how to organize assessment with multiple-choice questions within e-learning – what should be the number of the questions, how they should be ranked and ordered, how different parts of the test should be divided, how to provide feedback during online assessment etc.

We believe that the results of our research are beneficial both from the pedagogical and personalization perspectives of students’ online assessment. The instructors should not only carefully prepare the multiple-choice questions for such quizzes, but should carefully select the navigational strategy to be used during the assessment.

In this paper we demonstrated that indeed, flexible order navigation allows student to better distribute their efforts. However, providing additional information support to the student (like information about the difficulty of each question or about the time recommended for answering the current question) compensates for the limitation of strict order navigation through the test, thus allowing use the fixed order in order to get a benefit of computerized adaptive testing or immediate feedback adaptation mechanisms without negative impact caused by the fixed order navigation.
Our further research in this direction includes a more detailed analysis of the students' navigational paths and identification of strategies for passing the tests under different experiment conditions and with respect to students learning styles.

Acknowledgements

We would like to thank the students who participated in the courses and provided their valuable comments regarding the organization of the online tests. We are also thankful to the developers of the open-source LMS Sakai and its Mneme testing module, which has been used in our teaching and research. We would like to thank many people involved in the development of ProM tool, which we used for data analysis.

References


