Iterative student-based testing

of automated information-handling exercises

C.K. Ramaiah*, Mubarak Sulaiman** and A.J. Meadows***

*DESIDOC, Metcalfe House, Delhi, India

**Department of Library and Information Science, King Saud University, Riyadh, Saudi Arabia

***Department of Information and Library Studies, Loughborough University of Technology

The continuing rapid changes in how information is handled via computers means that training in computer-based information handling must itself undergo continuous modification. The investigations reported here examine the use of student feedback to improve presentation of hands-on exercises. Student responses appear to be reasonably consistent across both time and different student backgrounds. For the groups examined here, such factors as age and sex appear to have little effect on responses. The only significant factor is duration and extent of prior computer experience. Problems in hands-on exercises noted by the students therefore tend to have a continuing applicability, which suggests the value of iterative feedback in the provision of such exercises.

Introduction

Much laboratory teaching of information-handling involves students in evaluating information provided either online or via a computer package. A lecturer can help students carry out these tasks in a variety of ways. In particular, it is customary to provide students with hand-outs, and there is good evidence that such hand-outs are a valuable resource, especially for lower-ability students (see, for example Saloman, 1979). In many of these exercises, students are passive receivers of information, in the sense that they assess the information but do not change it. However, it is sometimes possible to use student feedback to change the original input. In this case, the users' mental models of the system can be employed to modify the user-interface set up by the original designer (see Moran, 1981). A number of experiments have been carried out in the Department of

Information and Library Studies at Loughborough University to examine how computer interfaces and instruction sheets used in teaching can be improved by student feedback. The present paper discusses examples of this work to help suggest both the factors to be taken into account and the sorts of changes involved. Our approach has been based on the concept of 'iterative usability testing', the value of which has recently been emphasized by Shneiderman (1993).

The teaching/learning model envisaged here is essentially teacher-focused, and supposes a teaching environment in which both hand-outs and practical assistance are provided (Joyce and Weil, 1980). However, part of our experiments involves student groups of differing sizes, so that the results also have some implication for group-focused teaching. The experimental approach is task-based, depending on the provision of well-structured tasks for students to perform at three levels. These levels are (1) simple (finding factual information from a hypertext database); (2) complex (evaluating the presentation of information via an electronic bulletin board); (3) design (setting up a HyperCard stack). For this work, teaching sessions with second-year students of information and library studies were employed. Experiments in categories (1) and (3) were carried out with the same groups of students, and are discussed first below.

Hypertext experiment

The exercises here were based on HyperCard. The first required the retrieval of factual information from a disk produced at Drexel University in the USA, which described the university and its facilities. The second involved a simple customization exercise to produce a daily diary. An initial pilot study was carried out with six students to give some estimate of the likely minimum teaching time for the exercise, and how it should be organized.

First study

The first full study involved 41 students who had no previous experience of hypertext. Two-thirds were 20 years old or less, four-fifths were female, and most had at least three years of computing experience. For teaching purposes, the students were randomly assigned to one of 10 groups, each consisting of three to four students. (The number per group was determined by the number of machines available in the laboratory. However, since more machines were on order for the following year, this offered the chance of experimenting with different-sized groups.) The student in the group who did the inputting was changed at specified intervals, and the groups were instructed to discuss each step. The groups were separated into two sets, each of five groups. The first set was given a short introduction to hypertext (specifically HyperCard), and then went hands-on immediately. The second set was exposed to a much more extensive demonstration and discussion of hypertext and HyperCard before going hands-on. All groups were given the same hand-outs, covering the background to hypertext, to the specifics of HyperCard, and to the exercises.

In the Drexel disk experiment, note was taken of the time required to retrieve the information and the number of false steps involved. Students were also asked to complete questionnaires regarding any problems they had encountered. Similarly, in the diary

customization exercise, note was taken of how far each group had progressed in the time available, and students were again asked to record any problems encountered.

No significant differences were found between the two sets in terms of time required. The same was true of most on-screen problems (for example, interpretation of menu options, or of icons). However, there were significant differences in terms of moving between screens, or (in the second exercise) of transferring items. For the Drexel disk, the less well-prepared set recorded twice as many problems in this category, and for the diary exercise, 50% more as compared with the better-prepared set. For students who had not encountered a mouse before, the physical handling of a mouse proved to be a major (though rapidly surmountable) problem, regardless of the set.

Second study

Changes were next made in the organization of the teaching and the hand-out materials in the light of student feedback from the first study. The customization exercise was simplified; the questions on information retrieval, which had previously been in random order, were re-ordered in terms of hardness (as perceived by the students); and a hand-out was prepared regarding use of the mouse. Students had indicated that three to four participants per group was too many, and that differences in computing experience were particularly irritating when discussing responses. For the repeat study, two thirds of the students were grouped into pairs with similar computing backgrounds; the remainder worked on their own. There had been adverse comments on the duration of both the long and short introductions to hypertext and HyperCard; in the revised exercise, all students were given the same intermediate-length introduction before going hands-on.

Set	Characteristics
A	Female; similar ages; familiar with Macintosh
B	Female; range of ages; not familiar with Macintosh
C	Male; similar ages; not familiar with Macintosh
D	Male and female; range of ages; familiar with Macintosh

Table 1: Groupings used in the second study

The teaching exercise was now repeated with the next cohort of second-year students – 45 altogether – who had again not had prior exposure to hypertext. In this case, a half were 20 years old or less and two-thirds were female – rather different from the first group – but, again, most had three years or more of computing experience. This time round, the students were divided into four sets. In contrast to the first study, these sets were selected to have certain group characteristics in common, as listed in Table 1. The first study had suggested that computing experience, particularly in the use of Apple Macintoshes, was the most important factor relating to the accuracy and speed with which the HyperCard exercises were carried out. Membership of these four sets was therefore determined on the basis of familiarity with Macintoshes, in order to test how significant this effect was. It was decided to investigate, at the same time, how useful the hand-outs were. Sets A and D were provided with detailed hand-outs, but were expected to work in the laboratory mainly on their own. Sets B and C were given condensed hand-outs, but could readily call on oral advice in the laboratory.

	Assessment of hand-outs	First Study	Second Study
	Medium-length hand-outs	Detailed hand-outs	Condensed hand-outs
Very useful	20%	35%	9%
Fairly useful	66%	57%	59%
Not very useful	14%	9%	32%

Table 2: Evaluation of hand-out material

In general, respondents in the second study indicated an appreciably higher level of satisfaction with the exercises than those in the first study. In terms of ability to cope with the exercises, there were significant differences between sets A and D, on the one hand, and sets B and C, on the other. Though, overall, problems in handling a mouse were reduced compared with the first study, the level was significantly higher for sets B and C than for A and D. There was a smaller, but still clear-cut difference between these sets in terms of problems with interpreting icons: sets B and C also found this more difficult. In addition, sets B and C encountered more problems in moving between screens and in using paint tools. The level of difference can be illustrated by the time taken by students to complete the customization exercise. They were allotted 90 minutes of class time for this (otherwise the project had to be completed in their own time). Almost all the students in set D (which contained those with most computer experience) completed the exercise in this time, as did over a third of the students in set A. Only a quarter of the students in sets B and C finished within the time limit. By way of contrast, the responses from all sets indicated that working in pairs was generally preferred to working alone.

Students were asked to evaluate the hand-out material they were given. Their responses to the question of content are given in Table 2.

Discussion of these responses with the students indicated that the provision of detailed hand-outs was regarded as desirable by all groups, regardless of computer experience or of the amount of other assistance that was available. The revised introduction to the course received a higher rating than either the long or short introductions used in the first study.

Electronic bulletin board experiment

An initial study was also made here in order to obtain a general view of how students reacted to different kinds of electronic bulletin board, and what problems they found in using them. To this end, 47 students were asked to access and report on four readily available bulletin boards. The method was to ask them to respond to a series of questions which could be answered via the information on each bulletin board. One of the bulletin boards was BUBL (the Bulletin Board for Librarians). It was decided to concentrate on this, paying particular attention to the section that contained practical exercises designed to provide training in the use of online services available over JANET. The question at issue was how well the students felt that this section of the bulletin board had been laid out for training purposes. In studying this question, the students obtained practice in the

evaluation of bulletin boards while experimenting with online training. The BUBL administrators had agreed to try and modify this section in the light of student feedback. Two successive studies were therefore planned. The first cohort of students would carry out their assessment of the bulletin board. The section concerned would be modified on the basis of their comments, and then reassessed by a second cohort of students.

First study
The first cohort consisted of 41 students, of whom some two-thirds were female and three-quarters under 25; a half had more than three years of computing experience, and a half had accessed electronic bulletin boards before. The students were divided into four sets, depending on their previous computing experience. Each set had four sessions devoted to bulletin boards. In the first two, the students worked in pairs and received a general introduction to electronic bulletin boards. In the next two, they had the opportunity to work individually on BUBL. During the latter sessions, the students answered (by electronic mail to the coordinator) a series of questions relating to the layout and design of the interface, ease of absorbing the contents, and ease of navigation. The students' responses indicated that they encountered most problems in the area of layout and design. For example, they noted a need for better spacing on-screen between instructions and other types of text, a need to use highlighting (or some similar method) for distinguishing between different types of heading, and a need to differentiate between levels of difficulty in the practical exercises (for example, basic, intermediate and advanced).

Second study
Changes were made to BUBL during the intervening summer vacation. Then the material was re-evaluated by a new cohort of 54 students. Their characteristics were very similar to those of the first group except that, owing to a change in the pattern of teaching, only some 15% had previously accessed bulletin boards. The teaching methods and questions asked were the same as in the previous year. A comparison of the comments from the first and second studies indicates that both cohorts were looking for similar factors when using the bulletin board. However, the respondents in the second study clearly preferred the revised layout and design of the interface. For example, the proportion of students who suggested the need for changes in the training section to make it more user-friendly fell from 84% to 27%. At the same time, student responses on the second time round also implied that the overall ease of use of the system had decreased. Further feedback was sought on this apparently contradictory result: it was found that the problem lay in a major increase in the amount of information available via BUBL from the first to the second year (this was reflected in the fact that the main menu contained 15 items in the first year and 25 in the second). In consequence, though the material was better presented on-screen in the second year, it often took more time and effort to track down any particular item of information. There were no systematic differences, in this case, relating to the students' computer experience.

Conclusions

The participants in these experiments formed a homogeneous group in the sense that they were second-year students in the same department. However, they had both a range of computing backgrounds and widely differing levels of interest in computer-based systems. From this viewpoint, they can be seen as a reasonably typical student group. The fact that

the types of problem they encountered in these experiments appeared to be common to all suggests that, regardless of experience and motivation, the whole group had a similar mental model of what they were trying to do. Indeed, from their responses, it was clear that they envisaged the computer as primarily an information provider. In terms of personal characteristics, there was no correlation between student responses and sex or age. There was a slight suggestion in the first investigation that students with a science background experienced rather fewer problems than those with a humanities background, but this was submerged by differences in computer background. Experience in using a Macintosh was clearly a factor, to the extent that it made sense to teach HyperCard to Macintosh users and non-Macintosh users separately. By way of contrast, using and evaluating the training section of BUBL was not significantly influenced by computer experience. The reason appears to be that, although some students had had prior experience with bulletin boards, none had previously been involved with online training. Prior computer experience in the first investigation actually did not affect the range of problems that a student encountered: it did affect the number of errors made in carrying out the exercises. It is useful here to draw a distinction between two types of error: 'mistakes' and 'slips' (Norman, 1983). Computer users begin with some kind of intention when they start using a system. A 'mistake' occurs when there is some error or deficiency in this intention. A different type of error occurs when the intention is correctly formulated, but something goes wrong in the attempt to implement it. Such an error can be referred to as a 'slip'. In the studies reported here, where conditions are fairly well defined, novice and experienced computer users tend to make similar 'mistakes', but the former make considerably more 'slips'. In working on these exercises, the student preference was clearly for operating in pairs, with both members having the same computer background. Whatever their level of computer experience, students preferred to have their hand-outs as detailed as possible. At the same time, they preferred the introductory lectures and demonstrations to be of moderate length, simply providing a framework which would allow them to appreciate the overall nature of an activity before they went hands-on.

Students who participated in the second experiments were not told what modifications had been introduced as a result of feedback from the first experiments. Their responses indicate that the modifications generally improved the acceptability of the teaching/ learning interface involved. So far as the present type of information-acquiring and information-evaluating activities are concerned, it appears that iterative feedback can improve computer-based teaching exercises. This result seems to apply to all three levels of task distinguished in our introduction: simple, complex and design.

References

Joyce, B.R. and Weil, M. (1980), *Models of Teaching*, Englewood Cliffs NJ, Prentice-Hall.

Moran, T.P. (1981), 'The command language grammar: a representative computing system', *International Journal of Man-Machine Studies*, 15, 3–50.

Norman, D.A. (1983), 'Design rules based on analysis of human error', *Communications of the ACM*, 26, 254–8.

Saloman, G. (1979), *Intuition of Media, Cognition and Learning*, San Francisco, Jossey-Bass.

Shneiderman, B. (ed) (1993), *Sparks of Innovation in Human-computer Interaction*, Norwood NJ, Ablex.