Does web-based automatic feedback improve learning behavior and examination grades? An experimental study

Reidar Kvadsheim, Mari Mehlen, and Harald Kittang
Faculty of Engineering, Oslo University College,
Oslo, Norway
Reidar.Kvadsheim@iu.hio.no Mari.Mehlen@iu.hio.no harald-oyen.kittang@iu.hio.no

Abstract – Recently, two of the authors reported that immediate, automatic feedback on assignments helped to increase study motivation as well as pass rate among engineering students attending an introductory course in statistics. These results were obtained in a non-experimental pilot study in 2005. The sources of error inherent in this research design led us to conclude that experimental research is needed to arrive at more reliable conclusions.

Following up this conclusion, we conducted a new study in 2006, using an experimental design assigning the participants randomly to one of two experimental conditions: The "web-supported" students received immediate, automatic feedback after having entered their responses on the assignments electronically. The "paper-supported" students received written feedback on their paper-based submissions several days later. The findings contradicted the results of the non-experimental pilot study: No significant differences between the groups were found with regard to final examination grades, study effort, and preferences with regard to the method for submitting assignments. These results demonstrate the importance of using an experimental approach in order to assess the usefulness of a web-based, automatic system for submitting assignments and obtaining feedback. To improve this system, future research should address the details of what the students do and how they think and feel in the learning process when they use or fail to use the system.

Index Terms – Automatic feedback, engineering education, research methods, learning behavior, computer-assisted learning, statistics

INTRODUCTION

A web-based learning-aid termed *FlexLearn* has been under development for three years at Oslo University College, Faculty of Engineering. The system may be implemented in all subjects in which students are required to solve problems with numerical solutions. The operation of the system is thus: The system presents the individual student with an assignment. The student works out his or her numerical solution and types it in the appropriate response box. The system then provides automatic and immediate feedback to the student on the correctness of the solution. In addition, if the solution is wrong, the student receives a response-dependent hint as to possible sources of the error.

If the student's solution is wrong, he or she is free to make another try at the same problem, the only change being that the numerical magnitudes have been assigned new values randomly. The student may repeat the number of trials ad libitum, each time receiving new feedback. In fact, this option is also available for students whose solutions are correct. Thus, they may keep on training in order to perfect their skills and improve their understanding. This system, whereby numerical values of problem parameters are assigned randomly, has the additional benefit of ensuring that all students are given different problems.

The FlexLearn method contrasts with the traditional paper-based method whereby students receive their assignments on paper, hand in their paper-based solutions, and get their solutions back from the teaching assistant after several days with brief indications of the correctness of the numerical solutions. The "paper" students were not given the chance to have another try.

There were four distinctive properties of FlexLearn that appeared to put it at a large advantage relative to the paper-based method and that, therefore, made it a highly attractive alternative:

- *Immediate feedback to the student.* This is expected to increase the student's motivation to work on problem solving compared to the situation when feedback is delayed for a lengthy period of time. Further, immediate feedback makes it easier for the student to cognitively associate the procedure he or she has used with the correctness of the result and, thus, to learn the method if it works and revise it if it does not.
- *The chance to have another try at a given problem (with new parameter values) whenever one wants.* This vastly expands the set of learning opportunities offered the student and, therefore, is expected to increase learning behavior.
- *All students receive different versions of a given problem (since parameter values are assigned randomly to each student).* This makes cheating less likely, since copying of the solutions worked out by others will fail. Accordingly, the student is expected to be under greater pressure to solve the problem him or herself. The "paper" students all receive the same version of the problem, so copying will be a more profitable shortcut.

- *Reduced tutor costs.* Once the system, including the assignments and their solution algorithms, has been developed, the application of the system as an "automatic tutor" requires much less tutor time than the running of a paper-based system requiring the tutor or the hired teaching assistants to check each assignment handed in by the students "manually". Since hundreds of students may be enrolled in a course, and the course be repeated year after year, the potential cost savings may be considerable.

In order to assess the usefulness of FlexLearn, a non-experimental pilot study was carried out by two of the authors in the spring semester of 2005. The participants were students attending an introductory course in statistics. All of them were required to use FlexLearn. The study was based on midterm questionnaire data and data on final examination grades. The questionnaire data were reported in [1] and the data on final examination grades in [2]. The major findings were that FlexLearn had promoted student study motivation and study behavior and helped to increase the pass rate in statistics. However, it was emphasized that systematic experimental research is needed to arrive at more reliable conclusions as to the usefulness of FlexLearn and the effectiveness of different ways of applying it.

In response to this need, a controlled, randomized experimental study was carried out with the students attending the corresponding introductory course in statistics in the spring semester of 2006. The study compares the "web-based" students who used the learning-aid FlexLearn with the traditional students who received their assignments on paper. The present paper reports the results of this improved study. We start by reviewing some relevant earlier research and then proceed to present and discuss our own data.

PREVIOUS RESEARCH

Immediate vs. delayed feedback

Theory and research suggest that immediate feedback at least under some conditions causes stronger training motivation and more training behavior, given that the feedback is regarded as helpful by the student and thus may serve as a reward. Research shows that the typical tendency is that the longer the delay of the reward the less attractive it is to the actor. As a consequence, the actor will be less likely to choose the action leading to this reward and more likely to opt for a given alternative course of action. More specifically, most evidence suggests that the subjective value assigned to the reward is discounted hyperbolically, e.g. [3], [4].

Now, as noted earlier, cognitive feedback such as, for example, information about the correctness of a submitted solution, is also expected to promote learning in other and more "mental" ways than by being an attractive event that triggers problem solving behavior that would not otherwise have occurred. Therefore, depending on the kind of cognitive feedback, the effect of the length of the delay interval may well be more complex than a hyperbolic decline in motivation or learning. In view of this, it is worth noting that most applied studies using classroom quizzes and verbal learning materials have found immediate feedback to be more effective than delayed feedback [5].)

Need for self-determination and competence

Based on evidence on a wide range of behaviors, there is by now a fairly large body of literature supporting the idea that people have a basic psychological need for autonomy and for competence. This supposedly innate, universal requirement makes people seek and respond favorably to conditions that provide for self-determination and offer the chance to demonstrate or acquire competence. FlexLearn offers the student unrestricted access to such conditions. Self-Determination Theory and concepts such as autonomy, self-efficacy, mastery motivation, and intrinsic motivation are among the constructs that are used to describe this kind of motivation and its consequences for behavior, cf. [6], [7].

Studies of automatic assessment

FlexLearn shares some properties with the AIM tutoring system for mathematics, cf. Sangwin [8].The major difference is that the AIM system assesses and offers rich cognitive feedback on the details of the student's manipulation of symbolic expressions, for example when he or she evaluates an integral or a differential, whereas FlexLearn primarily assesses students' numerical responses to assignments as true or false in addition to providing brief hints on likely sources of error. The greater complexity of the feedback makes AIM a more demanding and expensive system than FlexLearn. FlexLearn embodies a different and simpler strategy for promoting learning: It tries above all to engage the student's work motivation in several ways designed to increase the amount of self-initiated problem-solving behavior and task-relevant communication and information seeking.

IMPLEMENTATION OF FLEXLEARN AT OSLO UNIVERSITY COLLEGE, SPRING 2006

The students were given access to 6 assignments, one at a time at predetermined intervals. That is, each time a new theme was presented in a lecture, an appropriate assignment was released. The assignment came in three alternative versions corresponding to three levels of difficulty. The levels were denoted E, C, and A respectively, in line with the now common international grading scale, according to which E is the poorest passing grade and A is the best grade. The students were told to choose the version they preferred to solve. The data referring to these choices are not presented and discussed in the present article.

The students were free to submit their solutions whenever they wanted to in a response period of 14 days. At the end of this response period access to the assignment was closed. This "distributed" assignment schedule was intended to promote a focused and even level of activity over the semester. For half of the students, the assignments, the solutions submitted by the students, the feedbacks to the students, as well as individual and statistical information on the activities and results of the students were all administered

by means of the computer-based learning platform Classfronter, which is a generally available commercial product. Each assignment was "individualized" in the sense that the values of the parameters of the problem to be solved were determined randomly for each student. The other half of the students received the same assignments printed on paper in the traditional way, but the parameter values were the same for all students.

METHOD

A total of 246 students were registered as participants in the statistics course at the start of the semester. They were assigned into two approximately equally large experimental groups using a simple procedure which for practical purposes can be regarded as random. The official alphabetically ordered list of students was used. Student no. 1, 3, 5 etc. were all assigned to one group, whereas the remaining ones (i.e. student no. 2, 4, 6 etc.) were assigned to the other group. One of these two groups (the "FlexLearn" group) used FlexLearn for receiving and doing assignments and receiving feedback on solutions. The other group (the "paper" group) relied on the traditional paper procedure described earlier. For both groups three sets of data were collected:

- Two sets of questionnaire-based data relating to the learning process, gathered at respectively midterm and immediately after the written final examination.
- Data on the grades obtained in the written final examination.

TABLE 1
THE STUDENT POPULATION AT THE START OF THE COURSE AND THE RESPONSE RATE AT THREE POINTS OF DATA COLLECTION

Time / *Group*	START % (N)	Q1 % (n1)	Q2 % (n2)	Q1&Q2 % (n1&2)	EXAM % (n3)
FlexLearn	100 (106)	50.0 (53)	64.2 (68)	43.4 (46)	91.5 (97)
Paper	100 (130)	45.4 (59)	75.4 (98)	36.2 (47)	100.0 (130)

Explanations of abbreviations in Table 1.
START: The population at the beginning of the course (percent and number).
Q1: The response rate at midterm (questionnaire 1).
Q2: The response rate immediately after the written final examination (questionnaire 2).
Q1&Q2: The students that responded to both questionnaires.
EXAM: The students that took part in the written final examination.

Discussion. A relatively small number of students joined paper group after the initial establishment of the groups. Some of these had originally been assigned to the FlexLearn group but nevertheless handed in the assignments in paper format. Most had registered too late to take part in the original formation of the groups. This modest deviation from randomness represents a source of possible bias in group composition that should be taken into consideration in the interpretation of the results. Caution is also dictated by the relatively small rates of response to questionnaire 1 for both groups and to questionnaire 2 for the FlexLearn group.

RESULTS

1. The effect of FlexLearn on the final examination grades

The distribution of the final examination grades is summarized in Table 2.

TABLE 2
THE DISTRIBUTION OF GRADES AMONG STUDENTS USING FLEXLEARN AND STUDENTS USING PAPER FOR SUBMITTING ASSIGNMENTS

Grade	FlexLearn % (n)	Paper % (n)	p-value for difference between groups
A+B	22.7 (22)	24.6 (32)	0.367
C	19.6 (19)	20.8 (27)	0.413
D+E	28.9 (28)	33.1 (43)	0.248
F	28.9 (28)	21.5 (28)	0.105
Sum	100 (97)	100 (130)	

Discussion. Based on a 5% level of significance, Table 2 shows that there is no significant difference between the grades obtained by the students using FlexLearn and the grades obtained by the students using paper. This result contrasts with the result obtained in the 2005 non-experimental pilot study, which suggested that the FlexLearn students obtained better grades than the paper students. Given the expected advantages of FlexLearn relative to the paper method (cf. the Introduction), and in view of the results of the pilot study, this is a surprising finding that calls for an explanation. Two groups of possible explanations are:

Biased group composition (imperfect randomization).
Although the start population of students was split into a FlexLearn group and a paper group using a procedure assumed to be functionally equivalent to randomization, subsequent uncontrolled events changed the original composition of these groups somewhat, cf. the discussion of Table 1. We have no information leading us to suspect that this disturbance has biased the result in a particular direction, but we cannot exclude that this may be the case.

Unexpected negative effects of the use of technological learning aids.
Several possibilities may be imagined. Firstly, FlexLearn does not require the student to submit the chain of reasoning and calculations leading to a numerical result, but asks only for the result itself. Similarly, automatic feedback focuses on this result and ignores the underlying procedure. Together with the chance to repeat the submission for an unrestricted number of times, this may tempt students to engage in a relatively mindless process of trial and error with insufficient attention to the logical steps yielding the numerical

conclusion. In addition, when the immediate automatic feedback confirms that the result is correct, this may conceivably cause the student to overestimate his or her skills and prematurely stop further learning efforts within the relevant area of competence. We have so far no data on the kind of mental work that students engage in before they submit the solutions, but have some information on the amount of time they devote to problem solving, cf. Table 2 below.

2. The effect of FlexLearn on study effort (reported number of hours worked per week)

Both questionnaires asked the student to report the average number of hours per week he or she had spent on problem solving in addition to the scheduled contact hours. In the midterm questionnaire this average referred to the course period so far, whereas in the post-examination questionnaire the average number of hours per week referred only to the last month immediately preceding the written final examination. The distribution of responses at midterm and after the final examination is shown in Table 3 and Table 4 respectively.

TABLE 3
AVERAGE NUMBER OF HOURS PER WEEK SPENT ON PROBLEM SOLVING DURING THE SEMESTER (REPORTED AT MIDTERM)

Until midterm	FlexLearn	Paper	p-value for difference between the groups
Hours per week	1.75	2.25	0.047
SE	0.20	0.23	

TABLE 4
AVERAGE NUMBER OF HOURS PER WEEK SPENT ON PROBLEM SOLVING IN THE LAST MONTH OF STUDY (REPORTED IMMEDIATELY AFTER THE FINAL WRITTEN EXAMINATION)

The last month of study	FlexLearn	Paper	p-value for difference between the groups
Hours per week	3.04	3.39	0.128
SE	0.23	0.20	

Discussion. Apart from the familiar surge in work effort for both groups prior to the final examination, the two tables suggest that the FlexLearn students on average spend somewhat *less* time on problem solving than the paper students. The difference is consistent though moderate, but (given a required significance level of 0.05) it is significant only for the month immediately preceding the final written examination.

In any case, this tendency is opposite to the expected one, given the vastly expanded set of opportunities for learning offered by FlexLearn (cf. the Introduction). This could be a methodological artifact, given the somewhat flawed randomization procedure mentioned earlier, although we have no indications suggesting that this deficiency has produced the pattern of results in the tables. However, the pattern may also be due to a tendency for FlexLearn (in its current setup) to shortcut the learning process by discouraging the necessary logical thinking activities. Thus, the tables are consistent with one of the explanations offered for the unexpected similarity in final examination grades between the FlexLearn students and the paper students (cf. the discussion of Table 2).

3. Expressed student preferences with regard to method for submitting assignments

Immediately after the written examination, both FlexLearn students and paper students were asked to indicate which way of submitting assignments they would choose if given the chance. The results are summarized in Table 5.

TABLE 5
PREFERRED METHOD FOR SUBMITTING ASSIGNMENTS

Method used / *Method preferred*	FlexLearn % (n)	Paper % (n)	p-value for difference between groups
FlexLearn	66.2 (45)	38.8 38	0.256
Paper	33.8 (23)	61.2 (60)	
Sum	100 (68)	100 (98)	

Discussion. Table 5 shows that there is no significant difference in preferences between the FlexLearn students and the paper students in the following sense: A large majority in both groups prefer to submit the assignment in the same way as they have already practiced if given the choice. At the same time, a substantial minority in both groups would have chosen to use the submission method used by the other group.

This pattern runs counter to the expected motivating properties of FlexLearn noted in the Introduction. A flawed randomization procedure could have contributed to this (cf. earlier discussions). There may, however, also be aspects of some students' experiences of working with FlexLearn that weaken their desire to work with FlexLearn and thus counteract any experienced advantages. For example, it could be that these students are aware of their own inclination to engage in superficial trial-and-error behavior instead of the kind of learning behavior that promotes understanding. Such awareness may conceivably lead the students to prefer a submission method that do not offer the same possibilities for self-delusion and escape from actual learning work. So far, we have no data about this.

There is also a possibility that some students experience the use of FlexLearn as difficult or aversive for other reasons. Involving a more complex technology than pen and paper, it requires some explanation and training in the beginning. Although students in general seem to manage well, we cannot exclude the possibility that a measure of lasting resentment develops in some students. We lack information on such emotional reactions.

CONCLUSIONS

1. The results obtained in the experimental study contradict all the three main positive findings in the earlier pilot study:
   a. There was no significant difference in final examination grades between the students using the automatic web-based tutoring system FlexLearn and the traditional paper-based system for submitting assignments.
   b. Also, there was no significant difference in reported work effort between the two groups.
   c. Finally, there was no significant difference between the two groups with regard to how they preferred to submit the assignments: Both groups favored to approximately the same degree the method they had practiced in the course.

2. The results illustrate the crucial importance of the quality of the research method used. A non-experimental study, even when it involves control groups, may lead to conclusions that are later overturned in a better designed experimental study. At the same time, in the present study, there remains some uncertainty about the interpretation of the results due to certain remaining methodological imperfections.

3. In addition to underlining the importance of method, the results shift the attention away from general technology-focused research questions about the usefulness of the web-based automated tutoring system toward specific questions about how the system affects various details of what the students do and how they think and feel in the learning process.

4. Future research should address three challenges:
   a. Improve the design and execution of the randomization procedure used to compose the experimental groups. This will significantly contribute to making the results more trustworthy.
   b. Collect more data on details of what the students do and how they think and feel in the learning process, either by questionnaire or by interview.
   c. Modify the design of the FlexLearn system in order to increase the need for the students to reflect, reason logically, and seek understanding when they work on the assignments.

REFERENCES

1. Kvadsheim, R. and M. Mehlen, *Using FlexLearn to improve learning in mathematics and statistics: A web-based learning-aid based on student self-administration and automatic feedback.*, in *ICEE 2005 (INTERNATIONAL CONFERENCE ON ENGINEERING EDUCATION)*. 2005, Silesian University of Technology: Gliwice, Poland. p. 732-737.
2. Kvadsheim, R. and M. Mehlen, *Does immediate, automatic feedback help to increase study motivation and pass rate? Early experiences in statistics* in *9th International Conference on Engineering Education*. 2006: San Juan, Puerto Rico.
3. Frederick, S., G. Loewenstein, and T. O'Donoghue, *Time discounting and time preference: A critical review.* Journal of Economic Literature, 2002. **40**(2): p. 351-401.
4. Myerson, J., et al., *Discounting delayed and probabilistic rewards: Processes and traits.* Journal of Economic Psychology, 2003. **24**: p. 619-635.
5. Kulik, J.A. and C.-L.C. Kulik, *Timing of feedback and verbal learning.* Review of Educational Research, 1988. **58**(1): p. 79-97.
6. Amiot, C.E., P. Gaudreau, and C.M. Blanchard, *Self-determination, coping, and goal attainment in sport.* Journal of Sport & Exercise Psychology, 2004. **26**: p. 396-411.
7. Deci, E.L. and R.M. Ryan, *Handbook of self-determination research*. 2002, Rochester, NY: The University of Rochester Press. 470.
8. Sangwin, C., *Section A. Assessing mathematics automatically using computer algebra and the internet.* Teaching Mathematics and its Applications, 2004. **23**(1): p. 1-14.