Int. J. Engng Ed. Vol. 26, No. 3, pp. 642–654, 2010
Printed in Great Britain.

0949-149X/91 $3.00+0.00

Web-Based Automatic Feedback on Assignments in Statistics: How Can it Help Students Learn Statistics and Universities Reduce Costs?*

REIDAR KVADSHEIM, MARI MEHLEN, HARALD KITTANG
Faculty of Engineering, Oslo University College, Oslo, Norway . E-mail: Reidar.Kvadsheim@iu.hio.no Mari.Mehlen@iu.hio.no harald.kittang@gmail.com

A non-experimental study in 2005 suggested 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 at Oslo University College. In the follow-up study reported here we used 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 to 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 study: no significant differences between the groups were found with regard to final examination grades, study effort (with a certain qualification) and preferences with regard to the method for submitting answers. Running tutoring costs, however, were much lower for the web-supported than for the paper-supported students.Therefore, the present methodologically improved study strengthens the evidence that such learning support may help reduce running tutoring costs without significantly lowering final examination grades. Reinforcing this conclusion, certain remaining weaknesses in the experimental procedure open the possibility that the final examination grades of the paper-supported students have been inflated relative to those of the web-supported students. Moreover, questionnaire data and informal observations obtained during this experiment suggest that the tested web-based system of learning support can be combined with more traditional ways of promoting learning that may help increase learning with only a small increase in tutoring costs. These challenges with regard to the test methodology and the design of the learning-support system need to be addressed in new experiments.

Keywords: web-supported learning; automatic feedback; immediate feedback; tutoring costs; statistics

1. INTRODUCTION

E-LEARNING, IN GENERAL, and the use of automatic feedback on student submissions, in particular, may conceivably offer at least three kinds of benefits to students and educational institutions:

1) improved learning outcomes;
2) lower tutoring costs;
3) greater flexibility for the learner as well as for the institution.

These three concerns are all addressed by a web-based learning-support system termed *FlexLearn,* which has been under development for four 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:

1) the system presents the individual student with an assignment;
2) the student works out his or her numerical solution and types it in the appropriate response box;
3) the system then provides automatic and immediate feedback to the student as to whether the solution is correct or not.

If the solution is wrong, the student receives a response-dependent hint as to possible sources of the error or to possible remedies. The student now has the chance to make another attempt at solving the assigned problem. He or she may repeat the attempt if necessary, and may go on to do so until the answer is accepted as true or the student loses faith and abandons further attempts. At this stage, regardless of whether the assigned task was solved successfully or not, the student has the option to continue training on a new version of the same general problem. The 'same general' problem is here defined quite narrowly and numerically: the wording of the tasks is identical in the two trials, the only difference between them being that the

* Accepted 25 November 2009.

parameter values are dissimilar and that, if data are presented, they too differ between the trials.

For example, the student is presented with a set of measurement results from a certain normally distributed variable and is asked to calculate the expected value μ of the variable based on this set of data. If the submitted numerical answer to this question lies within some distance of tolerance from the correct exact value, the student receives an affirmation that the answer is true. If the student's response is outside this region, he or she will be advised to apply the estimator for the expected value. He or she may then submit a revised answer. This will immediately trigger a new feedback message whose content depends on whether the revised answer is within the distance of tolerance from the correct value. The procedure may be repeated as long as the student needs or wishes.

Now, if the student chooses to download a new version of the 'same' general task, he or she will be faced with an identical task formulation except that the set of measurement results will be different, and the normally distributed variable which has generated these data has an expected value and a standard deviation which differ from those identifying the normally distributed variable in the first version of the problem. Note that in order to encourage the further gain in skill perfection that may result from repeated practice, the student is free to continue to practice on the 'same' general kind of problem even if his or her submitted solution is correct.

Clearly, in cases such as the above example, we have no information for confidently predicting or reliably controlling what the student will actually learn as a result of feedback and repeated attempts. The distinction between 'understanding' and 'memorizing' is relevant here, as is the notion of 'transfer' of knowledge and skills to new contexts. These new contexts may differ from the context of learning to a greater or lesser extent, and successful transfer may involve various operations of thinking, abstraction and meta-cognition [1]. Thus, if the student in fact had chosen the right estimator but made a miscalculation due to carelessness or lack of calculation skills, the learning benefit may be one of increased attention and care during calculation and the development of more dependable calculation skills.

If, on the other hand, the student's error relates to his or her use of a wrong procedure, an inappropriate formula, or a lack of understanding of the statistical concepts of 'expected value', 'estimator', etc., the feedback may prompt the student to search for information explaining the content and use of these procedures, formulas or concepts in the course textbook. In the latter case, the resulting learning may be said to have a higher level of 'cognitive' content than in the former case, where learning of attentional control and calculation skills may be said to be of a more 'behavioural' or 'procedural' kind. Both kinds of learning are necessary and can be promoted by systems of learning support.

In the present example, the feedback message exhorts the student to apply a specific estimator and, therefore, seems to be more likely to address 'procedural' than understanding-based learning deficits. If the feedback message instead of pointing to a specific mathematical tool had referred the student to the relevant statistical problem area along with a reference to a chapter in the textbook, it might have done more to encourage reading and reasoning and thus help the student find, understand and apply the correct procedure in the situation.

In general, FlexLearn gives the teacher a lot of freedom to design tasks and feedback messages posing challenges at different levels of cognitive complexity. An important limitation, however, is that the system as designed and envisaged so far does not permit problem formulations whose solutions are symbolic expressions and not numbers. This limitation notwithstanding, the student trying to solve a problem will typically need to apply his/her understanding of the problem, reason logically and select and manipulate symbolic expressions to be able to arrive at a numerical answer. Thus, cognitive processes will be engaged, and cognitive faculties and skills of various kinds will be exercised and developed depending on the requirements of the task. Informal observations of students actively engaged in problem solving with FlexLearn lend some support to this. Since the students all downloaded different versions of a given problem (distinguished by random differences in parameter values and data), copying the solution worked out by a fellow student was not a viable method. Instead, the students often engaged in discussions of reasons for failure and the right way to proceed. Such social exchanges on problem solving may not necessarily undermine individual learning but may, on the contrary, stimulate relevant cognitive processes.

In the reported experiment the system required the student to solve all parts of the compulsory assignments correctly in order to fulfil the course requirement. If the student was unable to do so, despite the chance to make an unlimited number of attempts at each part of the assignment, he or she was required to contact the course tutor for help. Such consultations normally took about ten minutes. In this way, tutoring was driven by the need the student felt for learning assistance. In the experiment only five students among more than 200 availed themselves of such tutor assistance, thus adding little to the total costs of tutoring.

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 at the same problem. Still, these students were faced with the same course requirement as the FlexLearn students in terms of the number of assignments they needed to solve correctly. Also, the assignments were identical for the two groups, with one reservation, namely that for any given problem, the parameter values were the same for all 'paper' students, whereas they differed in a random manner among the FlexLearn students. For both groups of students, apart from the need to meet the course requirement, the students' submissions did not influence the final examination grade.

There are five distinctive properties of FlexLearn—and any similar system of automatic assessment—that appear to put it at an advantage relative to the paper-based method:

1) 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. Adding to these advantages, experience shows that the 'paper' students in many cases did not receive feedback at all, since they neglected to visit the tutor's office and pick up the returned submissions with the tutor's comments.
2) The chance to have another try at a given problem whenever one wants to. Thus, no limit has thus far been set on the number of trials. This vastly expands the set of training opportunities offered the student and, therefore, is expected to increase learning-promoting behaviour. Simultaneously, it greatly increases the number of chances to obtain feedback on attempted solutions, including hints that may stimulate thinking about the problem and facilitate search for information that is relevant to solving it.
3) Flexibility for student and tutor with regard to the where and when of studying and tutoring. Answers to problems may be submitted and feedback given everywhere and at any time provided the student has access to the Internet.
4) All students receive different versions of a given problem (since parameter values are assigned randomly to each student whenever she or he makes a try at this problem). 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.
5) 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 may be repeated year after year, the potential cost savings are considerable.

In sum, factors 1–4 are all hypothesized to stimulate the FlexLearn students to spend more time on problem solving, or spend this time more efficiently, than the paper-supported students. Such an increase in the amount and efficiency of problem solving behaviour is suggested to go along with an increase in the quantity and quality of the thinking involved in the behaviour and, as a consequence, in learning, which in turn will help the students obtain better final examination grades. This two-step hypothesis about the effect of student time use does not presuppose an assumption that an increase in the time devoted to problem solving, or in the efficiency of problem solving behaviour, automatically translates into strengthened learning and improved final examination grades. Such a desirable result will be the outcome of a complex mediating cognitive and behavioural process that involves the details of what the student does, thinks and feels, and how he or she interacts with the web-based learning aid and with other people, such as fellow students and the tutor. We hypothesize, then, that properties 1–4 of FlexLearn engage this complex mediating process in a beneficial way, so that the end result of the process in terms of final examination grades is at least as good as for the paper-based students.

Property 5 of FlexLearn entails the hypothesis that tutoring time (i.e. the time spent by faculty and by hired teaching assistants) will be less for the FlexLearn students than for the paper-supported students. Depending on the magnitude of this difference, the reduction in tutoring costs is potentially an important advantage of FlexLearn provided that the final examination grades are not worse than for the paper students.

We are aware that there is a growing wealth of web-based automatic assessment systems developed by commercial firms or academic institutions. However, information about their existence, properties, areas of application and effects are often not readily or freely available. What information there is typically focuses on the tool and rarely provides published research-based evidence on the usefulness of the tool in teaching and learning. Apart from this, the impact of these technological learning aids on student learning, motivation, work habits and tutoring costs in any concrete application will depend on the way the tool is configured and applied by the course tutor within the total learning environment of the course. FlexLearn makes no claim to being unique or better as a technological learning aid than other web-based automatic assessment systems. Our approach has instead been guided by the belief in the long-term advantage of sustaining a continuous process of research-based documentation and gradual improvement of the total learning-support system, i.e. the tool in combination with the way it is used by the student and by the tutor. It seems to us that such a systematic and persistent research-based development of tools and methods ought to play a larger role in engineering education.

As a background to this, an initial, comparative but non-experimental, pilot study of some effects of FlexLearn was carried out by two of the present authors in the spring semester of 2005. The participants were students attending an introductory course in statistics. About half of them were required to use FlexLearn, whereas the comparison group enjoyed the traditional paper-based support. The study was based on midterm questionnaire data and data on final examination grades. The questionnaire data were reported in [2] and the data on final examination grades in [3]. The major findings were that FlexLearn had promoted student study motivation and study behaviour and helped to increase the pass rate in statistics.

To reduce the sources of error inherent in this non-experimental study, a controlled, randomized experimental study was carried out with the students attending the corresponding introductory course in statistics in the spring semester of 2006. This improved study compared the 'web-based' students who used the learning-aid FlexLearn with the traditional students who received their assignments on paper. The findings were reviewed in [4]. The present paper discusses the main results of this study more thoroughly, adding some non-experimental observations that suggest possible explanations of the data. Taken together, these findings and explanations point to specific changes in the way we support student learning and to new experiments that test if these changes actually promote learning.

2. PREVIOUS RESEARCH RELEVANT TO ASSESSING THE MAJOR FUNCTIONAL PROPERTIES OF FLEXLEARN

2.1 Immediate vs. delayed feedback

Theory and research suggest that immediate feedback at least under some conditions causes stronger training motivation and better training behaviour, provided 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 with time, e.g. [5, 6]. Also, most applied studies using classroom quizzes and verbal learning materials have found immediate feedback to be more effective than delayed feedback, e.g. [7–9]. Some studies, however, report more nuanced conclusions, relating the impact of delay to the material to be learned and to the test method, e.g. [10, 11]. In general, all these studies refer to retention and reproduction of verbal information or to multiple-choice tests and not to the acquisition and application of skills at solving problems in statistics and mathematics. The question of the optimal timing of various kinds of feedback on various kinds of training tasks in statistics and mathematics remains an open one.

2.2 Need for self-determination and competence

Based on evidence of a wide range of behaviours, a fairly large body of literature supports the idea that people have a basic psychological need for autonomy and for competence. This supposedly innate, universal requirement makes people seek and respond favourably to conditions that provide for self-determination and offer the chance to demonstrate or acquire competence. FlexLearn gives the student freedom to choose where and when to train in problem solving and provides immediate feedback on the success or failure of each attempted solution, thus offering the student almost unrestricted access to conditions where personal choice can be exercised and increased competence and mastery be experienced. In contrast, paper-based learning support gives the student much less freedom and also greatly reduces and delays the feedback. Thus, two sources of training motivation will be weaker for paper students, a fact that is expected to lead to a reduction in training behaviour, i.e. to less time devoted to problem solving and/or to fewer attempts to solve problems. 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 behaviour, cf. [12, 13].

2.3 Studies of automatic assessment and automatic tutoring

There are many learning platforms today that offer automatic assessment and automatic feedback. Whereas they may share some functional properties, they tend to differ in a variety of details regarding, for example, scope and purpose, making summary comparison between systems difficult. For the same reason, it is also largely meaningless to compare the systems with regard to usefulness. Further, given the actual and potential importance of such tools, there is surprisingly little actual research on their usefulness, and few studies are experimental. We shall briefly comment on some reports on systems of learning support in mathematics.

FlexLearn shares some properties with the AIM tutoring system for mathematics, cf. Sangwin [14]. For example, the operation of both systems rests on the prior identification of:

1) a set of typical errors that students make when they solve a given problem;
2) a feedback message tailored to each of these typical errors.

When a student submits his or her solution to the problem, and if the solution is wrong, the system automatically identifies the typical error that best describes the submitted solution and immediately transmits the appropriate feedback to the student. When no typical error fits the student's incorrect solution, FlexLearn provides a more generally worded feedback message to the student.

There is, however, a major difference between the two systems with regard to strategy of learning support. The ambition of the AIM system is to promote student learning of advanced mathematics by assessing and offering cognitive feedback on the details of the student's manipulation of symbolic expressions, as, for example, when he or she evaluates an integral or a differential.

In contrast, FlexLearn embodies a different philosophy of learning support: It merely assesses the student's numerical response to an assignment as true or false in addition to providing a brief hint on likely sources of error and where the student may find helpful information. In this way, FlexLearn attempts to stimulate problem solving motivation and problem solving behaviour, as well as to promote self-reliance. To the extent that it succeeds in this, it may help to strengthen student understanding indirectly and not directly by providing feedback on the various logical steps in the problem solving process.

Moreover, as mentioned earlier, FlexLearn is being developed as an integral part of a more comprehensive learning-support system that includes non-electronic activities by students as well as tutors. The aim is to develop an inclusive learning environment whose various parts are adapted to each other in order to promote an optimally functioning overall system. The greater mathematical complexity of the feedback makes AIM a much more demanding and expensive system to develop and operate than FlexLearn. Costs are an important concern in many if not most institutions of engineering education, quite apart from the learning outcomes.

A very different approach from that represented by AIM, but also mathematically quite sophisticated, has been taken by Gwo-Jen Hwang and collaborators. They have developed a web-based system, ITED II, for testing and diagnosing student learning problems in mathematics based on historical assessment records [15].

Quite different from the automatic tutoring programs mentioned so far are two dissimilar programs developed and tested by Harskamp and Suhre [16]. Each of these programs embodies one of two distinct pedagogical approaches to learning support in mathematics, namely an 'instructional' approach and a 'constructivist' approach. Both programs rely on the use of certain mathematical tools and a system of hints that are supplied to the students either in response to student demand or in a pre-programmed way.

Whereas the electronic learning support systems mentioned above, each in its own distinctive way, provide a relatively narrow spectrum of support to the students in terms of the kinds of materials they offer the students, the nature of the exercises the students are invited to do and the kind of feedback that the students may obtain from the system when they respond, a different class of online instructional materials in statistics represent a much more comprehensive approach to learning support. For example, Larreamendy-Joerns and his colleagues [17] reviewed and compared six online statistics courses by extending the evaluative framework applied by George W. Cobb to assess statistics textbooks [18]. One of the many sets of criteria used by Larreamendy-Joerns and his colleagues describes properties of the electronic feedback that the students may obtain in response to their activities during the course. A crucial weakness relative to FlexLearn is that, for all the six courses, feedback is restricted to supplying correct answers to open-ended knowledge questions (for the students to compare with their own answers) and to the automatic grading of multiple-choice items. Another important limitation is that feedback is not automatic; the student needs to click an icon in order to get it.

3. EXPERIMENTAL TESTING OF FLEXLEARN AT OSLO UNIVERSITY COLLEGE, SPRING 2006

The students were given access to six assignments, one at a time at predetermined intervals. That is, each time a new theme was presented in a lecture, an assignment involving theme-relevant knowledge and skills 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. It turned out that there was no significant difference between the FlexLearn group and 'paper' group with regard to choice, and this variable therefore was not pursued any further in the present study.

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 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 Fronter, 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.

4. 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. Students 1, 3, 5 etc. were assigned to one group, whereas the remaining ones (i.e. students. 2, 4, 6 etc.) were assigned to the other group. One of these two groups used FlexLearn for receiving and doing assignments and receiving feedback on solutions. The other group relied on the traditional paper procedure described earlier. For both groups five 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;
- Informal tutor observations on the way the students responded to the group assignment and to the conditions of study support;
- Rough tutor estimates of tutoring costs for students in the FlexLearn group and for students in the traditional paper-supported group.

5. RESULTS

We obtained five sets of results from the study. These show, with appropriate nuances and reservations, that neither the FlexLearn group nor the paper group scored significantly better than the other one with regard to final examination grades. However, the FlexLearn group may have spent somewhat less time on problem solving in the course. Notably, both groups expressed a preference for the learning support system they themselves had practiced in the course. Most striking, though, is the large reduction in running tutor time costs for the FlexLearn group relative to the paper students. These results are presented in due order as follows.

First, to be able to assess these results, we note the somewhat flawed working of the randomization procedure and the variable but mostly low rates of response to the two questionnaires.

5.1 Effectiveness of randomization procedure

Under the conditions in this field experiment, three uncontrolled sources of error have to some extent distorted the group composition established by the original quasi-random assignment of students to the two groups. Two of these sources reflect student self-selection whereas one represents a conscious choice by the course tutor.

First, a relatively small number of students joined the paper group after the initial establishment of the groups. Some of these students had originally been assigned to the FlexLearn group but nevertheless handed in the assignments in paper format. Thus, they effectively changed their group membership on their own accord. Since the paper students got less feedback on their assignments than the FlexLearn group—and with a much longer delay—the students who switched to the paper group had nothing to gain in terms of feedback. Most of the students who joined the paper group late, though, had registered too late to take part in the original formation of the groups. For practical reasons, in order to avoid the extra burden of having to explain the FlexLearn procedure to half of the latecomers individually as they turned up, the course tutor decided to include them all in the paper group. Being late for course registration might conceivably be positively (or negatively) correlated with poor work habits, work motivation or talent, but we have no data relevant to assessing these variables.

Second, some of the registered members of the paper group also availed themselves of the FlexLearn support system by allying themselves with registered members of the FlexLearn group and using their passwords in order to gain access to the system. Accordingly, this subset of the paper group benefited from both learning-support systems and may, therefore, have obtained better final examination grades than would have been the case if they had abided by the original group composition. If so, the final examination grades of the paper group as a whole have been distorted in a positive direction, implying that the grades on average would have been lower if all group members had complied with the group composition rules.

Taken together, these deviations from randomness in group composition may favour either group when it comes to final examination grades, and the different impacts may also run in opposite directions and, therefore, cancel each other out. We have so far no reliable data on the size of these deviations from randomness. Overall, this group composition procedure is better than non-experimental procedures that do not attempt to randomize, but the procedure still leaves uncertainty regarding its lack of bias.

Future studies may try to reduce the three distortions as follows:

- Assign latecomers to the course to one of the two experimental groups in a random manner.
- Refuse to accept submissions on paper from students originally assigned to the FlexLearn group.
- For the FlexLearn group, reduce the number of

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)

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.

permitted submissions of response to a given problem from the unlimited number today to, for example, three. Thus, the group members would have only two chances to correct a mistake. Presumably, this would make them reluctant to allow paper students illegitimately to spend these chances by letting the paper students use their passwords to access FlexLearn.

5.2 Group sizes and response rates.

Table 1 gives the student response rates for the two questionnaires as well as the size of the groups at the start and the end of the course.

We note the variable but mostly relatively low rates of response to questionnaire 1 for both groups and to questionnaire 2 for the FlexLearn group. This dictates caution in interpreting the data relating to the learning process, such as, for example, number of hours worked and preference.

We now proceed to the substantive results.

6. STUDENTS USING FLEXLEARN AND STUDENTS RELYING ON PAPER SUBMISSION OF SOLUTIONS DID NOT SCORE DIFFERENTLY IN TERMS OF 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)	0.37
C	19.6 (19)	20.8 (27)	0.41
D+E	28.9 (28)	33.1 (43)	0.25
F	28.9 (28)	21.5 (28)	0.11
Sum	100 (97)	100 (130)	

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.

6.1 Discussion

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), this is a surprising finding that calls for an explanation.

6.2 Unintended disturbing effects of the experimental procedures

Flaws in the randomization procedure were discussed above. In addition, there is a certain part of the participants' knowledge whose impact for present purposes may be non-intended. Since this non-intended part of student knowledge was not be measured in the reported experiment, and since its impact on the difference between the FlexLearn group and the paper group cannot be regarded as random, the knowledge is a source of error of unknown magnitude. The participants knew they were taking part in an experiment; they had been informed of the nature of the two groups, and they needed to be aware of the group to which they belonged in order to fulfil their role as group members. It is conceivable that this knowledge made the students behave differently from what they would have done if they had submitted their solutions to the assignments without knowing that that they were taking part in an experiment and that another group used a different method of submission. For example, the experimental attention and the associated activities, discussions among the students, etc., may have made the students work more or in a better way, so that the final examination grades were better than they would otherwise have been. This is a complex set of cognitive and social factors—collectively termed a 'Hawthorne effect'—that possibly would have affected the FlexLearn group and the paper group about equally or, perhaps, have exerted a stronger effect on the 'privileged' group receiving the novel treatment, i.e. the FlexLearn group [19, 20].

It will be difficult to eliminate and even to measure Hawthorne effects or similar effects in a field setting, but interviews with the participants during or after the experiment may conceivably shed some light on attitudes and behaviour that reflect these kinds of effects.

6.3 Unexpected detrimental effects of the present version of FlexLearn

Several possibilities may be imagined. First, in its present version, 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 an unrestricted number of times, these work conditions 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. Moreover, when the immediate automatic feedback confirms that the result is correct, this 'success' may conceivably cause the student to overestimate his or her skills and prematurely stop further learning efforts within the relevant area of competence.

So far we have no data that directly describe the kind and amount of mental work that students engage in before they submit the solutions and, in particular, if FlexLearn students differ from paper students in this regard. Interviews with students during or after the experiment may possibly offer useful indications.

Indirectly, questionnaire data on the amount of time the students devoted to problem solving provide a hint that the FlexLearn students may indeed have been somewhat neglectful of the learning process relative to the paper students, (Tables 3 and 4)

6.4 FlexLearn students may have spent less time on problem solving than paper students

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 shows that the FlexLearn students at midterm reported having spent significantly *less* time on problem solving than the paper students. Table 4 suggests a tendency in the same direction in the last month of study, but this difference is not significant given a required significance level of 0.05. Together, the two tables confirm the familiar surge in work effort for both groups before the final examination.

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	0.05
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	0.13
SE	0.23	0.20	

6.5 Discussion

The tendency for FlexLearn students to spend less time on problem solving than the paper students is the opposite of the tendency we expected, given the vastly expanded set of opportunities for learning offered by FlexLearn. The result could be a methodological artifact caused by the somewhat flawed randomization procedure discussed earlier. However, the pattern may also be due to a tendency for FlexLearn (in its current version) to shortcut the learning process by offering the student an escape route from the necessary logical thinking activities.

6.6 Implications for teaching and research

Regardless of the explanation of the observed pattern, tutor impressions of student work habits indicate that the learning outcomes stimulated by FlexLearn can be improved by inexpensive modifications compelling the students to work more regularly and perform activities that may strengthen learning. At least three such measures readily present themselves:

- The FlexLearn students may be obliged to submit answers to more assignments than the six demanded in the reported experiment.

Actually, an increase to nine assignments was tried out in some groups of students in 2007, and the impact on student work habits, as recorded informally by the tutor, was positive. There is, however, a need to try this out experimentally, and measure the impact of an increase from, for example, six to nine assignments more rigorously.

- In order to increase the need for the students to think and understand before they submit the answers to an assignment, we may require the students to supplement their web-submissions of answers by handing in paper-based documentation of the reasoning and the calculations underlying these answers.

The downside to this is the additional costs that are incurred. Students as well as tutors will have to spend more time. This is an important consideration for all institutions with a tight budget for teaching. For the tutor, the acts of collecting documentation, assessing it and returning it to the students with comments all take time. With a large class of, say, 100 or more students the required extra tutor time may easily add up to many days if not weeks during the course. We must take into consideration, though, that the tutor need not read the documentation. In general, the extra tutoring costs will depend on whether the tutor chooses to inspect none, some, or all of this documentation. At one extreme, the tutor merely collects the documentation in order to dispose of it or leave it unread in a storeroom. In this case, the added costs for the tutor and the institution are virtually negligible (at least in the course of one semester). The question is, however, how student work on the documentation they submit will be affected if or when they get to know that the result of their labour will not be assessed.

The demand that the students submit paper-based documentation needs to be backed up by knowledge about how the students respond to this obligation.

Future experiments should therefore address five research questions:

1) Do students required to hand in paper-based documentation of the answers submitted electronically go about assignments in a way that produces better learning than students who are not obliged to do so?
2) Does the obligation to document the solution procedures lead to better learning even if the students receive no individual feedback on their submission of documentation (possibly apart from a confirmation that the submission has been received or accepted)?
3) Do different kinds and amounts of feedback on student submissions of documentation differ with regard to how much they promote student learning?
4) Do the learning benefits of having to submit documentation depend on additional requirements that the documentation conform to certain norms as to what constitutes an acceptable chain of reasoning in favour of the submitted answer to the assignment?
5) To what extent, if any, will student learning be adversely affected if the students are told or suspect that parts or all of the submitted documentation will not necessarily be read by the tutor but that there is a certain risk that it will be controlled? How does any negative impact depend on the magnitude of this risk and the consequences to the student if the quality of the documentation fails to pass the test?

- To further increase the need for students to understand the reasoning leading to an answer to a problem, the course tutor may impose a limit on the maximum number of web submissions of answers to any given assignment.

This will restrict the student's chances to hit the correct answer by guesswork and, hence, make thinking and understanding a more profitable way of dealing with an assignment. For example, whereas there is currently no upper limit on the number of web submissions of answers, this limit might be set to three.

Such a limit should not be imposed arbitrarily, however, but ought to be a reasonable choice in view of research-based knowledge about how such limits affect student work habits and learning.

Future experiments should therefore address the following question: how do student approaches to solving problems in statistics, and resultant student learning, depend on the number of chances they have to try out suggested answers electronically before making their final submission of the answers?

6.7 Most students in both the FlexLearn groups and the paper group expressed preference for the learning-support system practiced by their own group

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 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 practiced in the course, 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.

6.8 Discussion

This pattern runs counter to the expected motivating properties of FlexLearn noted in the Introduction. A flawed randomization procedure could have contributed to this. There may, however, also be aspects of some students' experiences of working with FlexLearn that weaken their desire to work with it and thus counteract any experienced

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	0.26
Paper	33.8 (23)	61.2 (60)	
Sum	100 (68)	100 (98)	

advantages. For example, it could be that these students are aware of their own inclination to engage in superficial trial-and-error behaviour instead of the kind of learning behaviour that promotes understanding. Such awareness may conceivably lead the students to prefer a submission method that does 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 find the use of FlexLearn 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.

In future studies, data on student reactions to the pedagogical and technical aspects of FlexLearn should be collected by means of questionnaires and/or interviews.

7. FLEXLEARN MAY HELP REDUCE THE RUNNING COSTS OF TUTORING CONSIDERABLY, BUT TOTAL SAVINGS TO THE INSTITUTION DEPEND ON POLICY AND TIME HORIZON

The total costs of *FlexLearn* to the university include investment expenditures of various kinds, maintenance costs and running expenses, also of various kinds. The investment costs include the costs of developing the web-based system for providing the students with individualized assignments, for student submission of numerical answers to the assignments and for tutor provision of immediate, automatic feedback on these submissions. The investment costs also include the working out of a library of assignments within the various areas of introductory statistics. The maintenance costs comprise, among other things, updating of hardware and software and tutor competence in response to normal technological developments. The running expenses include tutor costs relating to applying the system in any given semester. Table 6 shows an estimate of the latter costs in terms of tutor time, permitting a comparison between the running tutor expenses for the FlexLearn system of learning support and the traditional paper system.

7.1 Discussion

These tutor time cost estimates are 'pilot estimates' calculated on the basis of the time spent by the main course tutor on an arbitrarily selected small number (10) of assignments submitted by the FlexLearn students and the paper students. The time costs are likely to vary considerably from person to person and to depend on a host of other factors such as, for example, fatigue and boredom, which may increase with the number of assignments already handled. Also, these tutor costs do not include the time spent by the main course tutor for hiring, training and following up the tutor assistants needed to assess the submissions of the paper students.

Despite such serious methodological deficiencies, the difference in running tutor costs between the FlexLearn method and the paper method revealed by the present pilot estimate is so large that the potential savings to be obtained by means of FlexLearn should be explored further.

In the long run, with large numbers of students, with subjects and courses that are repeated from year to year, and when FlexLearn is used as a learning support in more than one subject, the running tutor costs are likely to be an increasingly dominant part of the total costs of the course and lead to increasing savings (in monetary units as well as a percentage of the costs of traditional paper-based learning support). This follows from the fact that the investment expenditures are relatively large in the beginning and tend to be much smaller later, when hardware, software, a library of assignments, etc. are available for repeated use. Hence, if the institution adopts a policy of promoting the use of FlexLearn as a learning-support tool in as many subjects as possible, considerable cost reductions should be possible.

Note also that the web-based FlexLearn system for presenting the students with individualized assignments, letting the students enter numerical answers, and for providing automatic feedback is a general one that may be applied to all subjects. Accordingly, when FlexLearn is adopted as a learning aid in a new subject, there is no need to develop the system from scratch. The major investment will be to adapt it to the specific requirements of the subject and the course by, among other things, developing a library of assignments and a set of response-contingent feedback messages for each assignment.

Further, maintenance expenses related to updat-

Table 6. Estimated running tutor time costs for FlexLearn support and traditional paper support

Method used ***Tutor time costs***	**FlexLearn**	**Paper**	**FlexLearn costs in percent of Paper costs**
Per assignment / 1 student	25 seconds	5 minutes	≈ 8 %
Per assignment / 200 students	1 hour 23 minutes 20 seconds	16 hours 40 minutes	
For a course with 9 assignments and 200 students	12 hours 30 minutes ≈ 1.5 workdays	150 hours ≈ 4 work weeks	

ing of software, upgrading of computers etc. will typically be moderate or zero and, moreover, should not be booked as FlexLearn costs exclusively but be distributed fairly among all those courses or activities that benefit from these computers and this software.

We have so far not estimated the investment expenditures and the maintenance costs since the initial conception of FlexLearn, nor have we made any estimate of the future investment and maintenance expenses that will accrue if one wants to adopt FlexLearn as a learning aid in a new course.

8. ACTIVITIES RELATING TO THE DEVELOPMENT AND TESTING OF FLEXLEARN HAVE HELPED TO INCREASE THE FACULTY'S INTEREST IN, AND ENGAGEMENT WITH, THE CHALLENGES OF TEACHING IN ENGINEERING EDUCATION

The Faculty of Engineering, Oslo University College, is under strong budget pressure to reduce student dropout and teaching costs while maintaining or improving academic standards. At the same time, the science, methods, and skills of 'soft' fields like teaching and learning have traditionally enjoyed little interest and even respect as an area of research-based competence, methods and tools. To raise questions regarding the behaviour of teachers, their methods and their beliefs about teaching and learning have largely been anathema.

Importantly, our efforts to develop and test an automatic device such as FlexLearn have forced us to specify concretely and precisely how we will support student learning behaviour and what the students need to do in order to benefit from this support. To make each of these detailed practical decisions we have had to explain why we think this way of supporting the students will help them learn in a better way than conceivable alternative ways of doing things, including the alternative of doing things in the traditional paper-based way. Further, we have had to strengthen our 'theoretical' pedagogical arguments by testing FlexLearn with our own students in order to see if it actually works in accordance with our pedagogical ideas. It is our impression that all the thinking, discussions, research activities, data, meetings, reports etc. that have taken place in this connection have helped to increase awareness in the faculty of the pedagogical challenges that confront engineering education. Crucially, by enforcing a more engineering-like approach to planning and execution in teaching, the application of modern technological tools such as computers and the internet to learning has made for rapport with staff members with a mindset shaped by their lives in engineering or the sciences.

In turn, this growing understanding has helped to increase resources for faculty projects aimed at improving teaching and learning in engineering education. A network of faculty members collaborating on such projects has formed and now makes up a recognized faculty pool of resources for initiatives, activities and exchange of ideas within this area. As an outgrowth of this, communication and project cooperation are expanding to include people outside the faculty who have interests and expertise in teaching and learning.

9. CONCLUSIONS

9.1 Almost no significant differences between groups

No significant differences were found between the students supported by FlexLearn and the students enjoying traditional paper-based support. This is true for the final examination grades obtained by the students and for their preferences with regard to the method for submitting answers. There was also no significant difference between the two groups with regard to study effort in the last month before the final examination, whereas the FlexLearn students at midterm reported somewhat less time spent on problem solving than the paper students.

9.2 Unsuccessful randomization may have favoured paper-supported students

Informal tutor observations suggest that some paper-supported students secretly resisted their group assignment by illicitly availing themselves of FlexLearn support in addition to the support allotted to the paper-supported students. This illegitimate and unrecognized advantage of some paper-supported students may have helped to boost their final examination grades above their 'true' paper-supported levels. Hence, the actual benefits of FlexLearn support relative to paper support in the reported experiment may conceivably include improved learning outcomes in addition to the reduction of tutoring costs. This methodological concern need to be addressed in new experiments.

9.3 Lower tutoring costs for the web-supported students

Running tutoring costs were much lower for the web-supported than for the paper-supported students. In consequence, for the institution, roughly equivalent learning outcomes (as measured by the final examination grades) combined with lower tutoring costs favours the FlexLearn method of learning support.

9.4 Possibilities for improvements of the web-based system of learning support

Also, informal observations of student work habits indicate that the learning outcomes stimulated by FlexLearn can be improved by inexpensive modifications compelling the students to work more regularly and perform activities that may strengthen learning. For example, the students may be obliged to submit electronically answers to more problems than demanded in the reported experiment. Further, in order to increase the pressure on the students to think and understand before they submit the answers, we may require them to supplement their web submissions of answers by handing in paper-based documentation of the reasoning and the calculations underlying these answers. The extra tutoring costs incurred by this depend on whether the tutor chooses to inspect none, some, or all of this documentation. The efficacy of such modifications of the learning-support system needs to be tested experimentally.

REFERENCES

1. J. D. Bransford, *How people learn: brain, mind, experience, and school*: National Academy Press. X, Washington, D.C. 2000, pp. 374 s.
2. R. Kvadsheim 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, *International Conference on Engineering Education 2005—ICEE 2005*. Silesian University of Technology: Gliwice, Poland. 2005, pp. 732–737.
3. R. Kvadsheim and M. Mehlen, Does Immediate, Automatic Feedback Help to Increase Study Motivation and Pass Rate? Early Experiences in Statistics, *9th International Conference on Engineering Education*. San Juan, Puerto Rico, 2006.
4. R. Kvadsheim, M. Mehlen and H. Kittang. Does Web-Based Automatic Feedback Improve Learning Behavior and Examination Grades? An experimental study. *International Conference on Engineering Education—ICEE 2007*. Coimbra, Portugal: International Network for Engineering Education and Research—INEER, 2007.
5. S. G. Frederick, G. Loewenstein and T. O'Donoghue, Time discounting and time preference: A critical review. *Journal of Economic Literature,* **40**(2), 2002, pp. 351–401.
6. Myerson, J., et al., Discounting delayed and probabilistic rewards: Processes and traits. *Journal of Economic Psychology*, **24**, 2003, pp. 619–635.
7. J. A. Kulik and C.-L. C. Kulik, Timing of feedback and verbal learning, *Review of Educational Research*, **58**(1), 1988, pp. 79–97.
8. R. E. Dihoff, G. M. Brosvic, M. L. Epstein and M. J. Cook, Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *Psychological Record*, **54**(2), 2004, pp. 207–231.
9. G. M. Brosvic and M. L. Epstein, Enhancing learning in the introductory course, *Psychological Record*, **57**(3), 2007, pp. 391–408.
10. G. M. Brosvic, M. L. Epstein and R. E. Dihoff, Retention of Esperanto is affected by delay-interval task and item closure: A partial resolution of the delay-retention effect, *Psychological Record*, **56**(4), 2006, pp. 597–615.
11. A. C. Butler, J. D. Karpicke, and H. L. Roediger, The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology—Applied*, **13**(4), 2007, pp. 273–281.
12. C. E. Amiot, P. Gaudreau, and C. M. Blanchard, Self-determination, coping, and goal attainment in sport. *Journal of Sport & Exercise Psychology*, **26**: 2004, pp. 396–411.
13. E. L. Deci and R. M. Ryan, *Handbook of Self-determination Research*. 2002: The University of Rochester Press, Rochester, NY. 2002, p. 470.
14. C. Sangwin, Section A. Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications*, **23**(1), 2004, pp. 1–14.
15. G.-J. Hwang, J. C. R. Tseng, and G.-H. Hwang, Diagnosing student learning problems based on historical assessment records. *Innovations in Education and Teaching International*, **45**(1), 2008, pp. 77–89.
16. E. G. Harskamp and C. J. M. Suhre, Improving mathematical problem solving: A computerized approach. *Computers in Human Behavior*, **22**: 2006, pp. 801–815.
17. J. Larreamendy-Joerns, G. Leinhardt and J. Corredor, Six online statistics courses: Examination and review, *American Statistician*, **59**(3), 2005, pp. 240–251.
18. G. W. Cobb, Introductory textbooks—A framework for evaluation—A comparison of 16 books, *Journal of the American Statistical Association,* **82**(397), 1987, pp. 321–339.
19. Wickström, G. and T. Bendix, The "Hawthorne effect"—what did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment & Health*, **26**(4), 2000, pp. 363–367.
20. De Amici, D., C. Klersy, F. Ramajoli, L. Brustia and P. Politi, Impact of the Hawthorne effect in a longitudinal clinical study: The case of anesthesia. *Controlled Clinical Trials*, **21**(2), 2000 pp. 103–114.

Reidar Kvadsheim is an associate professor at the Faculty of Engineering, Oslo University College, Norway. He received his Mag. Art. degree in sociology and Dr. Philos. degree in psychology from Oslo University. His reasearch interests include, among other things, behavioural choice, learning processes, and design of learning support systems.

Mari Mehlen is an associate professor at the Faculty of Engineering, Oslo University College, Norway. She received her Cand. Real. degree in physics from Oslo University in 1980. Her recearch interests include, among other things, renewable energy resources with a speciality in solar energy, statistical analysis of indoor climate problems, learning processes, and design of learning support systems.

Harald Kittang is currently working in the e-learning company Mintra (www.mintra.no) as a script writer and project manager. He recieved his master of engineering in mathematics from the Norwegian University of Science and Technology (NUST), in Trondheim, Norway. He has taught several courses in mathematics and statistics at NUST and Oslo University College, Norway. His mathematical research interests are in the field of stable homotopy theory, a branch of topology.