*International Journal of Engineering Education* Vol. 31, No. 2, pp. 505–520, 2015
Printed in Great Britain
0949-149X/91 $3.00+0.00


# Does Clicker Use Improve Exam Scores? A Controlled Randomized Experiment in a Bachelor-Level Course in Software Engineering

REIDAR KVADSHEIM, HÅREK HAUGERUD, HUGO L. HAMMER, ALFRED BRATTERUD and LAURENCE HABIB
Department of Computer Science, Oslo and Akershus University College of Applied Sciences, Oslo, Norway.
E-mail: reidar.kvadsheim@senior.hioa.no, harek.haugerud@hioa.no, hugo.hammer@hioa.no, alfred.bratterud@hioa.no, laurence.habib@hioa.no

This paper reports a study of clicker use within an undergraduate course in Operating Systems. It is based on a controlled, randomized experiment with a crossover design that measures learning outcomes by means of test questions at different levels of cognitive difficulty. The study aims to investigate whether clicker use primarily promotes superficial learning, whereby students reapply uncritically a previously seen solution in a new situation, or a more genuine learning whereby they analyze new situations and solve new problems. The results suggest that students attending clicker-based lectures obtain better exam scores than students attending corresponding traditional lectures in the same course. Moreover, the superior scores achieved by the students attending the clicker-based lectures were most pronounced for exam questions that required knowledge of the subject matter. The article concludes that clicker-supported lectures may be tried out helpfully in engineering education to promote learning. Advice is given as to how one may proceed.

**Keywords:** clickers; student response systems; web application; controlled, randomized experiment; crossover design; learning; knowledge of subject matter; bachelor course; software engineering; engineering education

## 1. Introduction

Immediacy and interactivity are recurring keywords in contemporary educational research (e.g. [1–3]). Another theme that shines through the Higher Education literature is that of budgets being under pressure due to a variety of factors, including reductions in state allowances and increased operating costs [4, 5]. In that context, there seems to be a growing interest in introducing technology into the classroom in order to allow for, among other things, interactivity, large class sizes, better learning, and increased student satisfaction.

Student active learning activities are at the core of most engineering educational programs, which typically have a strong focus on problem-based learning [6], guided inquiry module instruction [7], laboratory-based instruction [8], real-world experimental projects [9] and other learning-by-doing educational activities [10]. However, traditional engineering education has been described as having generally little focus on formative assessment in the classroom [11] and relies more heavily on summative assessment [12]. Although there are normally a number of arenas where engineering students receive feedback on their work, for example within the realm of laboratory-based instruction and project-based learning, it is less common to have mechanisms for feedback during lectures. Nevertheless, the recent literature points towards an increased focus on alternative forms of assessment in engineering education, such as continuous assessment [13]. In some cases, technology is used to support the processes of formative assessment. Amongst the many technologies available, clickers have been the object of some experimentation in engineering education in order to promote active learning in the classroom [14].

In the present study we shall focus on the following research questions:

1. Do students attending clicker-based lectures obtain better scores at their final examination than students attending lectures without clickers?
2. If students attending clicker-based lectures obtain better exam scores than students attending lectures without clickers, is the gain larger for exam questions that require genuine subject knowledge than for questions that can be answered with more superficial knowledge?
3. Do important administrative conditions of teaching and learning such as locating the lectures in the morning or after lunch, or early or later in the semester, affect the impact of clicker use on students' scores at their final examination?

To establish a context for the research questions and the way we address them, this article provides first an overview of what is meant by the term 'clickers',

* Accepted 1 September 2014.

then it gives a succinct review of the literature on clickers.

Clickers are increasingly popular learning aids that may differ in functional details and that may carry different names. The notion of a 'clicker system' generally refers to hardware along with a software application that gathers and stores student responses, and makes them available in aggregated and synthesized form, either graphically or in a tabular form to some or to all users. It may be noted that there is a range of designations for such systems, including 'personal response system', 'classroom response systems', 'student response systems', 'audience response systems', or 'electronic response systems', all of which refer to approximately the same concept.

The term 'clickers' is commonly used today to refer to dedicated handheld devices equipped with a set of buttons that can be pressed by users. They are typically used to get a large number of people to input, anonymously, their response to one or more questions by 'clicking' or pushing a button on their device, corresponding to what they consider to be the correct answer. While such devices have been used in classrooms since the 1960s [15], recent developments of clicker systems have resulted in the creation of 'clicker applications' that do not require dedicated clicker equipment, but that can be used via computers, mobile phones or tablets with a network connection. In such a case, what is called a 'clicker' is not a dedicated physical artifact, but rather a functionality that may be integrated in familiar electronic tools that the learner already uses for other purposes. In the pilot study reported in the present article the students used their mobile phones as platforms for the clicker applications.

The assumption underlying the use of clickers during lectures is that they enable or encourage the students and the teacher to carry out effective learning activities that would have been impossible during 'traditional' lectures or—perhaps more typically—to engage in a larger number of learning behaviors, or do so in a pedagogically better way or at a lower cost. Alternatively, the idea is that the instructor is able to teach a larger number of students at the same time than is possible in a traditional lecture format, without sacrificing student learning or increasing teaching costs.

In sum, the challenge for the teacher is to design and execute clicker-supported lectures in such a way that the materials, the setting, and the nature and timing of the teacher interventions really promote student learning or reduce teaching costs beyond what they are in the traditional case. An additional challenge is to measure the learning outcomes in a way that validly and reliably captures the learning goals guiding the planning and execution of the lectures, for example by constructing suitable tasks for the final examination.

Clicker systems have largely be seen as a modern technological and, therefore, exciting way of implementing more or less traditional and well-documented pedagogical methods, such as for example peer instruction [16, 17], cooperative learning [18], inquiry-based instruction [19], instruction based on discussion [20], instruction based on feedback [21], and instruction based on visualizations [22]. All of these important pedagogical tools may be, and indeed have been, applied in a traditional low-tech classroom as well as with the aid of clicker technology.

There is a large and varied body of research addressing the effects of clicker use on a range of factors including, among other things, exam grades, student learning, student learning behavior, and student learning motivation. It is typically concluded that the effect is beneficial [23, 24]. However, it has also been reported that the effect may be detrimental [15, 25].

Interestingly, one experimental study found that clicker use had a positive effect on exam scores but that the effect was limited to exam questions similar to those employed in the clicker sessions of the lectures [26]. In general, the research literature attributes any seemingly positive impacts of the clicker technology on student learning, engagement and satisfaction to the way in which it is used by the teacher and the learners, that is, to the pedagogical part of the total system [15, 27].

The study reported here aims to address the research questions through a randomized experiment with a crossover design. The article ends with suggestions as to how new studies may improve the evidence regarding the usefulness of clickers for learning, including how to adapt the methods of training and testing to make clicker use more effective and to make exam score a more valid measure of learning. For institutions that consider the adoption of clickers, some ideas are put forward as to how one may boost the prospects of successful clicker trials, and in a systematic, stepwise manner encourage the adoption of clicker-supported learning in even more courses using even more well-tested and better methods.

## 2. Method

### *2.1 The experiment*

The experimental study reported here is designed to test the hypothesis that the use of clickers may help students to improve their exam scores and, in particular, to see how our results relate to the discussed earlier findings, which were few and were seemingly contradictory. The participants were students attending an undergraduate 10 ECTS credits course in Operating Systems in the spring semester of 2012 at the Oslo and Akershus University College of Applied Sciences.

### 2.2 *Design of the study*

The lectures of the Operating Systems class at the college are normally given two hours a week for 14 weeks to a single group of second-year B.Sc. students. For the purpose of the present experimental study, these students were randomly divided into two groups, A and B, with the restriction that the groups should be equal with respect to gender composition, national background (local or foreign), and study path.

On the lecture day, there were two periods of lectures, one from 8.30 to 10.15 and one from 12.30 to 14.15. During the first lecture the students were introduced to the study and used the clickers for the first time. The next seven weeks the clickers were used in the lectures of group A, and the last six week the clickers were used in the lectures of group B. By letting both group, A and B, attend both clicker and traditional lectures, we may ensure that any differences between groups A and B are canceled out (under certain assumptions). Such an experiment is called a $2 \times 2$ crossover experiment [28].

The curriculum of the course was divided into four subjects. Subject W was lectured for the first three weeks, subject X for the next four weeks, subject Y for the next three weeks and subject Z for the last four weeks. The final written exam was divided into four parts, corresponding to the same subjects. Such a design made it possible to relate the final exam results of the two groups to the teaching method used in the lectures. Each of the four parts of the final exam consisted of several questions. Some of these questions were quite similar in form and content to the questions given in class using the clickers, merely requiring the student to remember and reproduce facts and/or repeat procedures that he or she used to solve the training questions relating to the subject area taught in a certain part of the course. This direct and 'shallow' relationship between training tasks and exam questions is denoted 'E' (meaning 'equal') in Table 1 below.

Other exam questions were also related to the questions discussed in class, but in a more subtle way, requiring a deeper level of understanding. The student still needs to apply the knowledge required to solve training questions pertaining to a certain part of the course, but the application involves more complex cognitive operations. He or she must think more abstractly, recognize instances of more general cases, assess the relevance of information, retrieve additional required knowledge, combine information, and arrive at logical conclusions. This more challenging cognitive relationship between training tasks and exam questions is denoted 'R' (meaning 'related') in Table 1.

Finally, the knowledge required to answer certain exam questions may have no identifiable relationship to the knowledge demanded by the training tasks in a given subject area. This state of 'no relationship' is marked 'NR' in Table 1.

The lecturer delivered the same lecture twice each day, once using clickers and once as a traditional lecture. This might have caused a difference in the quality; the second lecture might be better due to one extra rehearsal. We tried to balance out this effect by letting the lecture time of day be swapped half-way into the period of clicker usage as shown in Table 1. Then the students received lectures both before and after lunch in each period (provided that the attendance was the same in both parts of the period).

Table 1 gives an overview of the design of the experiment.

### 2.3 *Comments on Table 1*

#### 2.3.1 *Explanation of the relevance relationship notation in Table 1 (cf. the next to the last row)*

Consider, for example, the column for subject X. There are 18 clicker training questions that the instructor has asked and that the students have answered and discussed in the course of the four lessons devoted to this subject. Without going into the content and the form of the individual training questions, we note that they together are 'equal' to exam questions a and c, that is, E: a, c. Moreover, they are 'related' to the other three exam questions b, d, and e; thus they require deeper understanding of subject X, that is, R: b, d, e. Finally, there is no exam question in subject X, which has not somehow been covered by the training tasks in subject X, i.e. NR:—. And so on for the columns for other subjects.

For a more detailed explanation of each of the relationships R, E, and NR between training questions and exam questions, consider the example of the first row: To what extent does the student need to understand the subject matter of the clicker training questions in subject W and the solution methods used to arrive at the correct answers in order to successfully answer the various exam questions a, b, . . . , f in subject W?

R = Related: The student needs to understand the subject matter of the training tasks. He or she needs to see that this understanding is relevant to answering the exam question a, despite the dissimilarity in the form and content of this question relative to the training questions. He or she must then be able to

**Table 1.** Design of the experiment: groups, subjects taught, training questions, test questions, time

| *Time* / *Groups* | *Period* ➡ / *Time of day* ➡ | **1.1** *AM* | *PM* | **1.2** *AM* | *PM* | **2.1** *AM* | *PM* | **2.2** *AM* | *PM* | **Test** (= Final exam questions) *AM* |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | C: Clicker | **C** | | | **C** | | | | | Individual written exam. Same for groups A and B |
| | N: No clicker | | | | | | **N** | **N** | | |
| **B** | C: Clicker | | | | | **C** | | | **C** | |
| | N: No clicker | | **N** | **N** | | | | | | |
| ***Subjects taught*** | One subject taught in each period | **W** | | **X** | | **Y** | | **Z** | | **Subjects tested at exam: W, X, Y, Z** |
| ***Lectures*** | No. of lectures in each period | 4 | | 4 | | 3 | | 3 | | |
| C Lectures *(one pr. week, numbered)* | C training conditions: *Lecture, C questions, C-based learning proc.* | (C1= N1), C2, C3, C4 | | C5, C6, C7, C8 | | C9, C10, C11 | | C12, C13, C14 | | |
| N Lectures *(one pr. week, numbered)* | N training conditions: *Lecture, N learning proc.* | (C1=N1), N2, N3, N4 | | N5, N6, N7, N8 | | N9, N10, N11 | | N12, N13, N14 | | |
| Number of C questions | C questions in each subject ➡ | 6 | | 18 | | 10 | | 13 | | |
| Number of N questions: *Not specified* | N questions in each subject: *Not specified* | Trad. lectures. No plan for engaging students in question answering and discussion. Teacher demonstrates problem solving. | | | | | | | | |
| The relevance of the C training questions in each subject to answering the exam questions in this subject | Three levels of relevance distinguished: NR: *Not relevant to ...* E: *Equal to ...* R: *Relevant to ...* | Rel. of C questions to exam questions a,b,c,d,e,f in subj. W: NR: d, e, f E: - R: a, b, c | | Rel. of C questions to exam questions a,b,c,d,e in subj. X: NR: - E: a, c R: b, d, e | | Rel. of C questions to exam questions a,b,c,d,e,f, g,h in subj. Y: NR: e, f E: b, h R: a, c, d, g | | Rel. of C questions to exam questions a,b,c,d in subj. Z: NR: b, c E: None R: a, d | | *Exam questions in each subject :* *W: 6 questions* a,b,c,d,e,f *X: 5 questions* a,b,c,d,e *Y: 8 questions* a,b,c,d,e,f,g,h *Z: 4 questions* a,b,c,d |
| The relevance of the N quest.: *Not specified* | | | | | | | | | | |

mobilize and apply this knowledge in the new situation by logical arguments and calculations, in order to answer exam question a. And so on for the other subjects and exam questions.

E = Equal: The exam question is similar to (called 'equal' to) the relevant training question in subject area X. The student needs to recognize the relevant earlier training situation, remember the procedure used to solve the training task, and then merely to repeat this procedure to solve the current task a.

In this case, no knowledge of the subject matter is required, only an *episodic* memory of the training situation.

NR = Not related: The clicker-based training question is not relevant to solving the exam task. Accordingly, clicker training with this question is not expected to influence the correctness of the submitted answer to the exam question. The interpretation is similar for the other subjects and exam questions.

### *2.3.2 The groups and the experimental treatments*

The experimental treatment variable may assume two values, C and N, indicating which of the two conditions of learning that is relevant for the students taking part in a given lecture. C means that the students are expected to use clickers to answer questions posed by the instructor during the lecture, and the lecturer offers the students a dialogical process for doing so and for recording and following up the answers. N indicates that the students attend a traditional lecture without the use of clickers.

The randomized, controlled crossover design of this experiment served to ensure that all the members of groups A and B in the course of the experiment received both treatments, but in a reverse order: The members of A attended clicker lectures (C) in the first half of the experiment and traditional lectures (N) in the second half, whereas the members of B experienced the two kinds of lectures in the opposite order, thus relying on traditional lectures for the first half of the experiment and on clicker-based lectures for the second half.

These differences between the members of A and B with regard to the *location within the course* of the C and the N lectures and the *order* in which the group members experience the two kinds of lectures raises several concerns when it comes to comparing clicker lectures and traditional lectures with regard to effects on the learning or motivation of the students. In addition to these challenges following from the crossover design, we have the possible impact of the fact that the weekly clicker-based lecture and the corresponding traditional lecture always took place at two different times of the same day, one before lunch (AM) and the other after lunch (PM). By using the same lecturer and the same weekday for these two lectures instead of resorting to different lecturers and/or different weekdays, we have eliminated a variety of sources of variation.

However, this advantage does come at a price, since there may be differences in student learning depending on the time of the day, in the present experiment before and after lunch. A range of examples of time of day effects on learning and memory have been reported in the literature, suggesting that learning may be influenced by, among other things, the circadian rhythm in the students, cf. [29–32]. Similarly, the performance of the teacher in the lecture situation may conceivably also reflect circadian variations, although we have not found any published studies on this.

For each of the two lectures given by the instructor on a given day we have also registered whether it is the first or the second one. This is relevant for at least two reasons. First, the lecturer may presumably improve his/her teaching between the first and the second lecture due to the training and/or feedback received from the students during the first lecture. Second, the lecturer may, deliberately or not, change his behavior from the first to the second performance as a consequence of other factors such as boredom or fatigue.

In general, in order to argue persuasively for a causal effect and not merely to demonstrate a statistical association between teaching method and learning outcomes, we need the crossover design to be 'balanced' with regard to the students' exposure to the design-related factors that can affect learning. That is, all students in groups A and B should take part in the same number of clicker lectures (C) and traditional lectures (N), and the amount and nature of the distinguishing characteristics of these two kinds of lectures should remain the same throughout the experiment. Table 1 reveals that this balance is absent in the present experiment. Moreover, although certain aspects of the imbalance are readily visible in the table and can to some extent be taken into account, the significance of other aspects cannot be assessed for lack of information. Let us add a brief systematic explication of the imbalance:

*1. The number of treatments*

The number of clicker lectures was larger for group A than for group B, 8 vs. 7. In consequence, the number of traditional lectures was larger for group B than for group A, 7 vs. 6. (The first lecture in the course was a clicker-based training lecture for all students in groups A and B at the start of the course; that is, C1 = N1. The purpose of this lecture was to provide all students with a shared basis of understanding and motivation at the start of the experiment). Additionally, the number of clicker-based questions was also larger for group A than for group B, 24 vs. 23.

*2. The order of the treatments*

The members of group A went on from the initial clicker-based lecture to practice the same method in the next seven lectures before they had to revert to the traditional lecture mode for the remaining part of the course. In an almost total reversal of this pattern, the members of B also got an initial taste of the clicker-based method in the first lecture before they changed track to endure a sequence of seven non-clicker lectures, after which they returned to practicing the clicker method in the remaining lectures in the course.

*3. The timing of the treatments*

Whereas the design of the experiment is balanced with regard to time of day—that is, the number of lectures before and after lunch is the same—various other possibly relevant time-related factors are imbalanced, for example the amount of time between the lecture and the test of its effect on learning (in particular the final course exam).

*4. The learning-relevant differences between the C and the N treatments*
Some potentially important aspects of the treatments (i.e. learning-relevant events and processes in the C and the N lectures) were designed to vary over the experimental period or changed as a consequence of designed variations. These treatment variations and their effects (if any) on learning will escape detection if we limit our attention to the simple distinction between use and non-use of clickers; the variations concern the details of *how* the clickers and the traditional teaching techniques are applied. First, Table 1 shows that the subject matter taught in the lectures changed from one phase to the next one in the four treatment phases of the experiment. Second, the number of clicker questions asked by the lecturer and dealt with by the students and the instructor in the learning process differed among the subjects, as did the number of questions posed and dealt with per lecture. The last number can easily be calculated from the numbers in Table 1. Third, we do not know the amount of time that the students and the instructor spent on answering and discussing the different clicker questions; nor do we have data on the nature and the duration of the student and instructor activities that made up this process.

More generally, there has been no standardization of the character of the clicker questions over the project period, and no precisely defined standardization of the relationship between the tasks and the pedagogical process in the clicker-based lectures and the content and the form of the corresponding traditional lectures. In order to mitigate the consequences of this in the early stage of research, the series of clicker-based lectures was planned specifically for the present experiment on the basis of an existing series of traditional lectures tried out in the same course in earlier semesters; the traditional series of lectures was retained in the experiment and served as the basis of comparison. Moreover, since the new series of clicker-based weekly lectures and the parallel comparison series of traditional lectures were planned and conducted by the same teacher, and as the aim was to maintain a week-to-week correspondence between the knowledge contents of the clicker-based and the traditional lectures, we assumed this to offer some assurance of a certain ‘normal’ relationship between clicker-based and traditional teaching.

*5. The treatments actually received by the students*
The factors 1–4 are all potentially important influences on student learning due to the crossover design of the present experiment; for us they are sources of possible treatment imbalances between groups A and B that may add to or detract from the impact of the planned pedagogical differences between the treatments of students in the groups A and B.

An additional, but related, source of uncertainty regarding the interpretation of the outcome of the experiment is the oversimplified and primitive nature of the binary distinction between clicker-based and traditional lectures (C vs. N). The experimental design outlined in Table 1 is a *plan* indicating which treatment (C or N) is to be administered to the members of groups A and B, and at what time during the experiment. The treatment data used in the reported results below refer exclusively to these planned, binary defined treatments. However, the treatments *actually received* by each student may have been different from the planned ones and, moreover, deviate from each other in other ways that may be unforeseen and their impact on learning may not have been considered. First, the success of each planned treatment depended on whether the student in fact attended the relevant lecture as opposed to staying away. Second, the student may have chosen to attend the alternative lecture instead of the planned one, or even to take part in both of the lectures, the planned one as well as the alternative one. Third, and most important, the learning-relevant contents of any treatment lecture include more than ‘mere presence’ during the lecture, particularly during the clicker-based lectures, where the students are expected to be active problem solvers by answering questions and taking part in the subsequent discussions. Thus, the actual treatment needs to be described in more precise aspects and in the nuanced terms of the kinds and amounts of behavior and other process aspects.

*2.4 Test of the effect of the experimental treatment.*
Two effect tests based on final examination scores are alluded to in Table 1:

1. Do students who have attended clicker-based lectures in a given subject obtain better final examination scores in this subject than students who have attended traditional lectures in the same subject? If so, how large is the clicker effect?
2. Does the clicker effect (if any) depend on the relationship between the final exam questions testing the students’ knowledge in a certain subject and the training questions posed by the lecturer in order to promote student learning in the subject?

*Comment:* Based on the earlier review of the literature, the hypothesis here is that the clicker-based training advantage effect is larger for exam questions that presuppose genuine subject knowledge than for questions that require none or only superficial subject knowledge. Accordingly, we expect the

**Table 2.** Mean exam scores for groups A and B on exam questions in subjects that were taught in periods 1.1, 1.2 and periods 2.1, 2.2

| | **Periods 1.1, 1.2/Subjects W, X** | **Periods 2.1, 2.2/Subjects Y, Z** |
|---|---|---|
| Group A<br>*(No. students: 27)* | Mean exam score: 76.2 (C)<br>*Stdev exam score: 23.4* | Mean exam score: 66.4 (N)<br>*Stdev exam score: 26.2* |
| Group B<br>*(No. students: 17)* | Mean exam score: 74.4 (N)<br>*Stdev exam score: 24.4* | Mean exam score: 77.2 (C)<br>*Stdev exam score: 23.4* |

clicker effect to be larger for exam questions that are 'related to' a certain earlier clicker-based training question than for exam questions that are 'equal to' the clicker-based training question. For those exam questions that do not require any knowledge of the subject (the 'equal' category), the notion of a clicker advantage effect is more complex. There might be a clicker advantage effect, though a 'fake' one, if the clicker-supported student has failed to acquire real subject knowledge but luckily has picked up the answer to a training question that happens to re-emerge as an exam question. Or, conceivably, there could be some general effect of clicker-based training that benefits all exam questions, if the training somehow inspires the students to change the way they study or to devote more time to learning.

## 3. Results and discussion

Table 2 shows mean exam scores, on a scale from 0 to 100, with standard deviations for the four main parts of the experiment.

### *3.1 Comment on Table 2*

The label (C) indicates that the group in question had attended clicker lectures, whereas (N) means that the group had attended a traditional lecture where clickers were not used. For example, group A took part in clicker lectures in periods 1.1 and 1.2, while group B did not. The table shows as a summary result that the mean exam score in a given curriculum is largest for those students who attended clicker lectures in the periods when the curriculum was taught. This result—viewed in isolation—suggests a 'yes' answer to the first part of the main research question 1 stated earlier: 'Did the students attending the clicker-based lectures in a given subject area obtain better scores at the final examination in this area than the students attending the parallel 'control' version of the lectures without clickers?'

The 2 × 2 crossover experiment can be analyzed using the Student's $t$-test. See Jones and Kenward [28] chapter 2 for details. The results are summarized in Table 3.

### *3.2 Comment on Table 3*

The clicker effect is tested by computing for each student the difference in exam score for the topics when the student used the clicker in the lectures and for the topics when the student did not use the clicker. We see from Table 3 that the effect is significant with $p$-value 0.031, meaning that exam scores related to the clicker lectures are better than for the traditional lectures. A 95% confidence interval for the clicker effect is [1.2, 24.3] with the mean value of 12.7, which corresponds to an improvement in exam results of more than one grade (e.g. from E to D or B to A).

**Table 3.** 2 × 2 crossover trial. Result

| | **Degrees of freedom** | **Tobs** | **p-value** |
|---|---|---|---|
| Period effect | 42 | 1.2439 | 0.220 |
| Carry-over effect | 42 | –0.6502 | 0.519 |
| Clicker effect | 42 | 2.2315 | 0.031 |

By carry-over effect we mean the difference in exam score as a result of the order in which the students attended the traditional and clicker lectures. In many crossover experiments we expect to observe such an effect if, for example, the students' use of clickers in the first part of the experiment produce a change in learning behavior that endures in the second part of the experiment. We see from Table 3 that for this experiment we do not observe any significant carry-over effect.

By period effect we mean the difference in exam scores for periods 1 and 2 in the experimental set up, for example because the curricula are different in the two periods (i.e. W, X vs. Y, Z). We see from Table 3 that we do not observe any significant period effect for this experiment.

In sum, the results so far are compatible with the hypothesis that the use of clickers during lectures did help the students obtain better exam scores. This hypothesis is a speculative and tentative conclusion, however, that says nothing about why and how clicker use influenced exam scores. A possible brief answer to these questions is:

1. that the clickers helped the students learn more than they would have done without clickers, and
2. that the resulting increase in knowledge enabled the students to give better answers to the exam questions than they would otherwise have done.

We may assess the plausibility of this answer by inspecting more detailed data about the exam

**Table 4.** Descriptive data. Equal

| | Periods 1.1, 1.2/Subjects W, X | Periods 2.1, 2.2/Subjects Y, Z |
|---|---|---|
| Group A<br>*(No. students: 27)* | Mean exam score: 78.7<br>*Stdev exam score: 25.3* | Mean exam score: 62.6<br>*Stdev exam score: 35.4* |
| Group B<br>*(No. students: 17)* | Mean exam score: 80.0<br>*Stdev exam score: 23.3* | Mean exam score: 70.6<br>*Stdev exam score: 31.7* |

**Table 5.** Descriptive data. Related

| | Periods 1.1, 1.2/Subjects W, X | Periods 2.1, 2.2/Subjects Y, Z |
|---|---|---|
| Group A<br>*(No. students: 27)* | Mean exam score: 76.6<br>*Stdev exam score: 22.6* | Mean exam score: 64.7<br>*Stdev exam score: 27.1* |
| Group B<br>*(No. students: 17)* | Mean exam score: 75.9<br>*Stdev exam score: 23.4* | Mean exam score: 77.8<br>*Stdev exam score: 22.2* |

**Table 6.** Descriptive data. Not related

| | Periods 1.1, 1.2/Subjects W, X | Periods 2.1, 2.2/Subjects Y, Z |
|---|---|---|
| Group A<br>*(No. students: 27)* | Mean exam score: 74.8<br>*Stdev exam score: 33.4* | Mean exam score: 68.3<br>*Stdev exam score: 32.2* |
| Group B<br>*(No. students: 17)* | Mean exam score: 67.1<br>*Stdev exam score: 39.2* | Mean exam score: 76.6<br>*Stdev exam score: 32.2* |

scores, relating the data to the research and theory reviewed in the Introduction to the present paper.

First, we note (cf. Table 1) that the exam questions differ with regard to the kind and extent of the knowledge that the students need to have and be able to use in order to answer correctly, the knowledge referring to what was taught and trained for in the relevant lectures. Some exam questions, denoted NR ('Not Relevant') do not presuppose subject knowledge taught in the course. In these cases clicker training during the lectures does not seem helpful. To answer other questions marked E ('Equal'), the student needs superficial understanding of the subject. It is sufficient if he or she remembers a solution procedure that was used in a certain training task and has the good sense or luck to feel that the exam question is of the same kind so that the same procedure can be used. In fact, in the present study, it was enough for the student to remember the answer to the clicker question and not the procedure originally needed to arrive at the answer. In both of these cases clicker training during the relevant lectures may conceivably have aided some kind of learning (cf. Carnaghan and Webb [26]), but to a lesser extent than in those cases when the exam questions are labeled R ('Related'). In order to provide correct answers to such questions the student needs real understanding of the subject matter, so that he or she realizes that the subject knowledge is applicable to the exam question and is able to pick out the knowledge elements that are needed and to apply them to the current task. In support of this logic, we refer to the research and the extensive literature review by Mayer et al. [24], which suggest that clickers work by stimulating deep cognitive processing of the subject content taught and applied in the lectures.

Now, consider Tables 4–6, which have the same general form as the summary overview of exam scores and standard deviations in Table 2, with the important difference that the exam scores and standard deviations are calculated and displayed separately for each of the three categories of exam questions: Equal, Relevant, and Not Related. Visual inspection of the three tables does not suggest any major differences among the results for the three categories.

Table 7 provides a helpful statistical overview of these data by showing $p$-values and 95% confidence intervals for the clicker effect for the three categories of exam questions.

Only for the question category 'Related' is there a statistically significant effect of clicker use on exam score (p-value = 0.045) on a 5% significance level, with a mean improvement in score of 13.8. For the categories of 'Not related' and 'Equal', the clicker effects are not significant (with p-values of 0.099 and 0.49, and mean improvements in exam score of 16 and 6.7 respectively).

**Table 7.** 2 × 2 Crossover trial. Clicker effects

| | Degrees of freedom | Tobs | p-value | 95% confidence interval |
|---|---|---|---|---|
| Equal | 42 | 0.69 | 0.49 | [–12.8 , 26.2] |
| Related | 42 | 2.066 | 0.045 | [0.3 , 27.2] |
| Not related | 42 | 1.687 | 0.099 | [–3.1 , 35.1] |

One might speculate on the difference between the results for these last two categories and the apparently smaller difference between the categories of 'Related' and 'Not related', although we do not have strong evidence that the improvements in exam scores pertain only to one or two particular question categories or that they are significantly larger for some categories. In this situation, it gives food for thought that the only significant clicker effect has been observed for the exam question category 'Related', which requires the students to selectively retrieve and apply genuine knowledge of the subject matter. This is consistent with the earlier research and theory suggesting that the benefit of clicker use above all has to do with stimulation of student behavior and thinking geared to understanding the arguments and steps in the solution of problems. Consequently, it is not surprising to find indications that clickers are of less help in tasks that do not require much reasoning but merely that students are able to remember and repeat a procedure or answer they have practiced earlier. At this point, it is worth noting that the data in Table 7, suggesting that clicker effect is smallest for those exam questions that are similar to the training questions, is contrary to the finding by Carnaghan and Webb [26], who reported the largest clicker effect in this case.

Let us add a note of caution. In the analysis and discussion of the results we have for convenience we assumed that the design of the experiment is unbiased in the sense that it does not include irrelevant factors that favor clicker lectures over traditional lectures or the other way around, when it comes to influencing final examination scores. We have also, however, pointed out a number of reasons suggesting that this assumption is unlikely to be fully satisfied.

Table 8, below, illustrates one of these reasons. The table shows two aspects of observed student activity: 'Actual' student attendance in the traditional lectures and 'actual' student clicker activity in the clicker-based lectures (as distinct from the 'planned' number of lectures to be attended and the 'planned' number of questions to be answered by the students as given in Table 1). In Table 8 attendance at the traditional lectures is summarized by the mean number of lectures actually attended by the students in a given period, and the amount of clicker activity in the clicker lectures is summarized by the mean number of clicker questions actually answered by the students during a given period.

*Comment on Table 8.* We see that we have seasonal differences in both of these measures of student activity. The number of traditional lectures attended by the students in group B in periods 1.1, 1.2 in the winter season is notably larger than the number of traditional lectures attended by the students in group A in periods 2.1, 2.2 in the spring season, i.e. 6.9 vs. 4.6. On the other hand, there was a similarly unbalanced lecture plan of eight lectures in the winter season as opposed to only six in the spring season cf. the design of the trial in Table 1. As for student clicker activity in the clicker lectures, the imbalance between winter and spring is similar: the mean number of clicker questions answered by the students in group A in the winter season is notably larger than the number answered by group B in the spring season, 26.7 vs. 22.2. There is a similar, but smaller, difference between the total number of clicker-based training questions in the two seasons, 24 vs. 23. We refrain from speculating on if and how these activity data can be expected to influence clicker effects on exam scores.

We end the discussion of the results with two brief comments.

1. The test used to look for a potential carry-over effect (cf. Table 3) is rather weak, meaning that we could have a carry-over effect even though the test is non-significant. The test for clicker effect is done under the assumption that we do not have any carry-over effect.
2. The students are unlikely to have started their participation in the experiment with an unbiased attitude regarding the benefits of clickers vs. traditional teaching-focused lectures. Informal feedback from the students suggests that they expected clickers to be a useful tool to stimulate learning. It this is true, it might bring about a positive clicker effect via a mechanism of 'self-fulfilling prophecy'.

### *3.3 Supplementary note on effect tests and categorization of knowledge*

Final examination scores are used as test instruments because it is assumed that student differences

**Table 8.** Student lecture attendance and student clicker activity

| | **Periods 1.1, 1.2/Subjects W, X** | **Periods 2.1, 2.2/Subjects Y, Z** |
|---|---|---|
| Group A<br>*(No. students: 27)* | Mean no. clicks: 26.7<br>*Stdev no. clicks: 6.3* | Mean no. lectures: 4.6<br>*Stdev no. lectures: 2.3* |
| Group B<br>*(No. students: 17)* | Mean no. lectures: 6.9<br>*Stdev no. lectures: 2.7* | Mean no. clicks: 22.2<br>*Stdev no. clicks: 6.6* |

in such scores, under certain conditions, validly and reliably reflect differences in the amount of student knowledge of the subject matter. The distinction between examination questions that are respectively 'related to' and 'equal to' questions used for clicker-based training serves to delimit one important condition for the validity of exam score as a measure of student knowledge of the subject matter at the end of the course. The reason for the introduction of this distinction is that we are faced with the challenge of measuring competence in an applied educational setting, with important research-relevant constraints, and not in a laboratory.

The simple classification of exam questions used in this study tries to capture one important factor, correspondingly roughly to the commonsense distinction between 'easy' and 'difficult' exam tasks. An easy task that is correctly answered will mislead the person assessing the answer to credit the student with more subject knowledge than he or she really has and, accordingly, assigns the student a too high grade. More optimistically, the measured impact could turn out to be the positive one of improving the student's answers on difficult exam questions and thus obtain a well-deserved higher grade. In fact, both could be the case; that is, the clicker-based lectures could turn out to improve true subject knowledge as well as merely apparent knowledge.

### *3.4 Supplementary note on the reliability and validity of the test instrument (i.e. the final examination score as a measure of knowledge of the subject matter taught)*

The students' knowledge at the end of the course was measured by assessing their submitted answers to the various questions constituting the final written exam. The person conducting the assessment did not know the identity of the students and, hence, also did not know which students attended clicker-based lectures and which students did not. Also, the exam tasks in the software engineering subject 'Operating Systems' are such that the correct answers typically enjoy agreement among experts making this assessment. Accordingly, the reliability of this measure (both in terms of the stability of repeated assessments made by the same teacher and in terms of the concordance between different teachers assessing the same answer) is believed to be high, although it has not been measured. We also regard the expert assignment of exam questions in Operating Systems to the categories of 'equal to', 'related to' and 'not relevant' as a simple and reliable judgment.

The *validity* of exam score as a measure of student subject knowledge at the end of the course is a more complex issue. This is due to the fact that we cannot know whether an exam question that is 'equal' to a clicker-based training question merely incites the student to copy the recalled answer to the training question or if the student accesses his knowledge of the subject matter and actively processes the relevant parts of it to arrive at the answer to the question. For those exam questions that are 'related to' earlier clicker-based training questions, such uncertainty does not exist. Hence, this subset of exam questions makes for valid exam scores. In consequence, total exam score may overrate the subject knowledge possessed by some students in the non-clicker group. The doubtful validity of the set of 'equal to' questions implies that total exam score in our study could be biased in favor of the non-clicker group, thereby exaggerating the subject knowledge demonstrated by these students relative to the knowledge shown by the students attending the clicker-supported lectures. While this is true in principle, the pattern of student answers indicates that such a bias may be small or non-existent in the present study.

In order to see if there was a difference in validity between the two categories of exam questions in our data, we have calculated the difference in exam scores between clicker students and traditional students separately for each of the categories of exam questions, in addition to computing the overall difference in exam scores between the two student groups. The results are in the expected direction, although the data does not permit definite conclusions.

### *3.5 Summary of the discussion*

In sum, the results suggest that the controlled randomized experiment offers improved, but still limited, evidence that the use of clickers to engage students in problem solving and discussions during lectures can stimulate long-term student learning and help the students obtain better test scores in the final examination.

The improvements in the evidence are due to two aspects of the design of the study:

1. the study is a controlled, randomized trial with a crossover design, and
2. measurements of learning outcomes relies on different kinds of questions testing knowledge with either much substantive content or consisting merely in an episodic memory of the correct answer to a particular clicker-based training question.

The first design aspect, the controlled, randomized crossover methodology, helps to reduce, and assess statistically, the risk that the observed effect of clicker use is really brought about by clicker use and not by other factors that co-vary with it and independently influence learning. The second design aspect, the application of multiple dissimilar effect measurements focusing on different kinds of learning outcomes, helps to see if the observed improvement in learning (if any) is selective in the expected direction. Does the improvement concern exclusively, or mostly, measurement items that probe the level of subject knowledge presumably generated through the clicker-related learning activities?

The uncertainty surrounding the size and the interpretation of the results is related to several aspects of the design and the execution of the study:

- The relatively small number of participants and their possible biased initial attitudes and beliefs, along with the practical difficulties associated with delimiting the set of participants at the start of the study and dividing this set into two subsets by a random procedure.
- The small number of exam questions at different levels of knowledge in the various subject categories. This limits the possibilities for studying the possible effect of clicker use on the acquisition of knowledge at different levels of depth.
- The lack of sufficiently detailed information about the content of the traditional lectures and the absence of a known, precise relationship between these lectures and the clicker-based lectures. This hinders comparison between the two approaches and makes identification of likely learning-relevant differences between the two pedagogical processes difficult.
- The execution of the study in full accordance with the planned design is hard to achieve in the somewhat untidy setting of university teaching. We have not assessed how far we have succeeded in this. In particular, we have not explored data on the actual activities of the individual students in the various lectures, for example: Did he or she attend the lecture? Did he or she answer the clicker-based questions? Did he or she take part in the discussions? If yes, how?
- The inherent limitations of controlled, randomized experimental crossover studies of human learning, including difficulties in determining and interpreting carry-over effects and period effects. Also, there is no subset of participants who took part in a totally traditional version of the course, with no use of clickers. Similarly, a subset of participants experiencing a fully clicker-based version of the course is not available for comparison.

In future studies, there will be a need to establish a clear correspondence between the contents and activities of the traditional lectures and those of the alternative clicker-based learning sessions in order to identify learning-relevant similarities and differences between them.

## 4. Conclusions

Although the study was carried out in the subject of Operating Systems, its potential usefulness is not limited to this subject, nor to software engineering or to engineering studies in general. The only immediately evident limitation is that the subject matter taught must be a structured body of knowledge (in terms of concepts, relationships, tools, and procedures) that enjoys a sufficiently high degree of consensus among recognized experts, so that actionable learning objectives can be formulated and their attainment measured with validity and reliability.

With this in mind, the reported study suggests the following.

1. Clicker-supported lectures may helpfully be tried out in one or more subjects as a tool for achieving significant student gains in exam scores as well as in knowledge of the subject matter. (In the present study the average gain in score corresponded to an advance from one level to the next one on the usual six-level scale from A through F, from for example E to D, or C to B.)
2. To prepare for effective clicker-supported training and thereby to strengthen the chances of success, the institution and the teachers concerned should check the current quality of the framework for teaching and learning with regard to the course (or courses) concerned:
   (a) Are the learning objectives stated clearly in terms of actions (tasks) that the students are expected to be able to carry out? Can the students' ability to do so be assessed?
   (b) Do the current procedures for preparing and carrying out the final examination (or any other tests for measuring the learning outcome) ensure that the exam tasks validly test the students' ability to perform the actions defining the learning objectives?

   If the first or both of these questions cannot unreservedly be answered with a 'yes', there is a need to identify the reason and eliminate it. Otherwise, the teacher(s) responsible for the course will lack the direction and the quality assurance needed to prepare and carry out an effective clicker-aided course plan.
3. It may be prudent for an institution, in view of the limited and uncertain available research-based knowledge, to start with just one, or a few courses, in order to gain experience, and then proceed gradually. Trusted colleagues that report positive results, and that can provide practical advice, can increase the willingness and the ability of others to make a try.

4. Institutions that consider the possibility of implementing a new, clicker-supported version of an existing course and measuring its effectiveness, may find the description and the discussion of the methodology of the present study a helpful source of ideas and challenges.

## References


1. D. J. Skiba and A. J. Barton, Adapting your teaching to accommodate the net generation of learners, *Online Journal of Issues in Nursing*, **11**(2), 2006, pp. 1–7.
2. S. Kelly and L. Fall, An investigation of computer-mediated instructional immediacy in online education: a comparison of graduate and undergraduate students' motivation to learn, *Journal of Advertising Education*, **15**(1), 2011, pp. 44–51.
3. B.-C. Seet and T.-T. Goh, Exploring the affordance and acceptance of an e-reader device as a collaborative learning system, *Electronic Library*, **30**(4), 2012, pp. 516–542.
4. E. D. Capaldi, Budget cuts and educational quality, *Academe*, **97**(6), 2011, pp. 10–13.
5. H. S. Bienen, The financial future of research universities, *Social Research*, **79**(3), 2012, pp. 631–634.
6. R. Da Silva et al., PBL and B-learning for civil engineering students in a transportation course, *Journal of Professional Issues in Engineering Education & Practice*, **138**(4), 2012, pp. 305–313.
7. C.-S. Cho et al., Enhancing education of construction materials course using guided inquiry modules instruction, *Journal of Professional Issues in Engineering Education & Practice*, **139**(1), 2013, pp. 27–32.
8. J. Bernhard, Insightful learning in the laboratory: some experiences from 10 years of designing and using conceptual labs, *European Journal of Engineering Education*, **35**(3), 2010, pp. 271–287.
9. T. E. Baldock and H. Chanson, Undergraduate teaching of ideal and real fluid flows: the value of real-world experimental projects, *European Journal of Engineering Education*, **31**(6), 2006, pp. 729–739.
10. L. Bot et al., 'Learning by Doing': a teaching method for active learning in scientific graduate education, *European Journal of Engineering Education*, **30**(1), 2005, pp. 105–119.
11. L. McDowell, S. White and H. C. Davis, Changing assessment practice in engineering: how can understanding lecturer perspectives help? *European Journal of Engineering Education*, **29**(2), 2004, pp. 173–181.
12. N. W. Heap, K. L. Kear and C. C. Bissell, An overview of ICT-based assessment for engineering education, *European Journal of Engineering Education*, **29**(2), 2004, pp. 241–250.
13. S. Myllymäki, Incorporation of continuous student assessment into lectures in engineering education, *European Journal of Engineering Education*, **38**(4), 2013, pp. 385–393.
14. B. Ray and R. Bhaskaran, Integrating simulation into the engineering curriculum: a case study, *International Journal of Mechanical Engineering Education*, **41**(3), 2013, pp. 269–280.
15. M. E. Lantz, The use of 'Clickers' in the classroom: Teaching innovation or merely an amusing novelty? *Computers in Human Behavior*, **26**(4), 2010, pp. 556–561.
16. C. H. Crouch and E. Mazur, Peer instruction: ten years of experience and results, *American Journal of Physics*, **69**(9), 2001, pp. 970–977.
17. K. R. Wenzel and D. E. Watkins, Instruction based on peer interaction, in *Handbook of Research on Learning and Instruction*, R. E. Mayer and P. A. Alexander (Eds), 2011, Routledge, New York, pp. 322–343.
18. R. E. Slavin, Instruction based on cooperative learning, in *Handbook of Research on Learning and Instruction*, R. E. Mayer and P. A. Alexander (Eds), 2011, Routledge, New York, pp. 344–360.
19. S. M. M. Loyens and R. M. J. P. Rikers, Instruction based on inquiry, in *Handbook of Research on Learning and Instruction*, R. E. Mayer and P. A. Alexander (Eds), 2011, Routledge, New York, pp. 361–381.
20. K. P. Murphy, I. A. G. Wilkinson and A. O. Soter, Instruction based on discussion, in *Handbook of Research on Learning and Instruction*, R. E. Mayer and P. A. Alexander (eds) 2011, Routledge, New York, pp. 382–407.
21. J. Hattie and M. Gan, *Instruction based on feedback*, in *Handbook of research on learning and instruction*, R.E. Mayer and P.A. Alexander (eds) 2011, Routledge, New York, pp. 249–271.
22. R. E. Mayer, *Instruction based on visualizations*, in *Handbook of Research on Learning and Instruction*, R. E. Mayer (Ed.), 2011, Routledge, New York, pp. 427–445.
23. H. M. McGowan and B. K. Gunderson, A randomized experiment exploring how certain features of clicker use effect undergraduate students' engagement and learning in statistics, *Technology Innovations in Statistics Education*, **4**(1), 2010.
24. R. E. Mayer et al., Clickers in college classrooms: fostering learning with questioning methods in large lecture classes, *Contemporary Educational Psychology*, **34**(1), 2009, pp. 51–57.
25. L. A. Van Dijk, G. C. Van der Berg and H. Van Keulen, Interactive lectures in engineering education, *European Journal of Engineering Education*, **26**(1), 2001, pp. 15–28.
26. C. Carnaghan and A. Webb, *Investigating the Effects of Group Response Systems on Student Satisfaction, Learning and Engagement in Accounting Education*. Available at SSRN: http://ssrn.com/abstract=959370 or http://dx.doi.org/10.2139/ssrn.959370, 2006.
27. S. M. Keough, Clickers in the classroom: a review and a replication, *Journal of Management Education*, **36**(6), 2012, pp. 822–847.
28. B. Jones and M. G. Kenward, *Design and Analysis of Cross-Over Trials*, 2nd edn, 2003, Chapman & Hall, London.
29. M. J. Anderson et al., Individual differences in the effect of time of day on long-term memory access, *The American Journal of Psychology*, **104**(2), 1991, pp. 241–255.
30. C. P. May, L. Hasher, and N. Foong, Implicit memory, age, and time of day: paradoxical priming effects, *Psychological Science*, **16**(2), 2005, pp. 96–100.
31. C. P. May, L. Hasher and E. R. Stoltzfus, Optimal time of day and the magnitude of age differences in memory, *Psychological Science*, **4**(5), 1993, pp. 326–330.
32. T. V. Petros, B. E. Beckwith and M. Anderson, Individual differences in the effects of time of day and passage difficulty on prose memory in adults, *British Journal of Psychology*, **81**(1), 1990, pp. 63–72.


## Appendices

### *A. Description of the clicker app*

For the purpose of this study we have developed a mobile web application (web-app) in-house, i.e. involving both students and members of the academic staff. The motivation behind developing in-house was to be able to continuously adapt the app to the needs of the users. The rationale behind choosing to perform this study with an app rather than a traditional clicker was as follows:

- The cost of traditional clickers was not insignificant, especially considering that this study was the first of a much wider research project, intending to include several thousands of students, and that the exact number of participants was unknown.
- The next stage of the research is supposed to be carried out in several countries, which made it impractical to purchase physical clickers.
- A preliminary survey amongst the students of the faculty revealed that 95% of the students had an Internet connected device with them during classes and lectures, albeit on a number of different platforms including iPhone, Android, Mac and Windows.
- A solution based on a web-app would solve the problems described as far as cost, availability and scalability are concerned.
- Traditional clickers allow only for a limited set of question types, whereas a web-app allows for a much wider range of uses.

After it became clear that traditional clickers were not to be used for this study, the question of what platform to choose was an easy one, since there were no apparent advantages to developing separate platform-specific applications. Web was the obvious multi-platform solution to the problem, i.e. targeting both mobile devices and laptops, both of which were used in classrooms.

The functionalities of the app are as follows.

- It was built on a client–server model where the clients (mobile devices and laptops) send their answers to a server, which stores the data and provides a web page allowing a visual presentation of the results as diagrams.
- The client had a very simple user interface, based on a structure of channels, whereby each channel can include an indefinite number of questions.
- All users (both teachers and students) had the possibility to create channels and create questions within them. This functionality was created to give an incentive for students to use the app for extra-curricular purposes.

*B. Supplementary note on the method of the experiment*

The experiment, using a cross-over design with the random assignment of students to clicker and non-clicker (i.e. 'traditional') conditions, was run for the first time as a regular bachelor-level course that the students need to pass to earn their degree. Clicker questions and clicker use had to be planned and coordinated with suitably adapted lecture parts. All of this, in turn, had to be coordinated with the traditional non-clicker version of the course, so that the two versions of the course could be run in parallel and could prepare the students for exactly the same exam. This required a lot of work for the single person in charge of planning and running both of the course versions.

Despite this effort, there is no simple full correspondence between the two versions of the course in this pilot experiment. Still, the curriculum, the requirements, and the final exam tasks are identical, and the timing of the lectures and their subject areas also overlap almost completely. Also, the questions that the instructor asked the students, or the task solutions that he demonstrated, and the concepts that he explained, were also mostly the same. However, not every question asked, solution demonstrated, or explanation given to one group also turned up in some form for the other group. Or, if it occurred, it was for example offered as an immediate and unasked for answer by the instructor to a question that he had posed himself as part of the lecture. This complex relationship between the clicker (C) and non-clicker (N) versions of the course manifests itself in Table 1 in the absence of specifications of the training tasks as well as their relevance for the final exam tasks in the case of the non-clicker version. The result is that any difference (or similarity) in the final examination scores of the students between the two course versions may be explained in more than one way: Does it have to do with clicker use or is it due to other differences or similarities between the versions? Precisely how?

As has been explained and exemplified in the quasi-experiment by Mayer et al. [24], it is difficult or meaningless to conceive of a 'clicker effect' as such; the clicker system and the way it is applied are but a particular means for realizing, for example, a certain pattern of communication between the instructor and the students, a pattern that may also be realized by other, non-electronic means. Accordingly, the clicker method not only needs to be described accurately but must be contrasted with a precisely defined alternative instructional design.

*C. Examples of clicker-based training questions and exam questions*

The following example questions are taken from subject Y: 'Threads, synchronizing and internal memory'.

*i.* Example questions from the category *equal*

The first of these clicker questions is literally equal to the corresponding exam question, (b). In the second example there are some small differences and the corresponding exam question, (h), is a bit more detailed. However, the essence of the questions is equal and so is the essence of the answers. In the first example the question concerns a quite complex mechanism that most students find hard to fully understand, but in order to be able to answer the exam question correctly, the student can base the answer on just remembering the answer from the clicker test.

Clicker question:
What will happen if a process fails to signal a semaphore after a critical section?

Clicker response alternatives:
1. The process will be stopped by the operating system.
2. The operating system will, after a while, signal the semaphore.
3. No other processes will thereafter be able to enter the critical section.
4. The critical section will no longer be critical.
5. Two other processes will be able to enter the critical section simultaneously.

Exam question:
(b) What will happen if a process fails to signal a semaphore after a critical section? Explain briefly.

Exam answer:
Then the semaphore will remain locked and other processes (if they behave properly) will not be able to enter the critical section.

Answer number 3 of the above clicker response alternatives is the correct one. The exam answer is a bit more elaborate, but a student answering exactly the same as the given clicker answer would be given full score at the exam.

Clicker question:
A program is running and uses 256 MB of internal memory. Where is the page table located when the process is running?

Clicker response alternatives:
1. The entire table is in the MMU.
2. Some in the MMU, the rest in RAM.
3. The entire table is in RAM.
4. Some in RAM, the rest in swap.

Exam question:
(h) A program is running and uses 512 MB of internal memory. Where in the hardware is the page table located when the process is running? Do if necessary include locations where parts of it may reside. Explain briefly.

Exam answer:
Assuming a page size of 4 KB there will be approximately 128 thousand entries in the page table. That is far too much to be stored within the CPU in the MMU and the TLB and some parts of the page table will reside in RAM. Parts of it may also reside in other parts of the cache than TLB, in L2 and L3.

Answer number 2 of the above clicker answers is the correct one. The exam answer is more elaborate, but a student answering exactly the same as the given clicker answer would be given full score at the exam.

*ii.* Example questions from the category *related*

The first of these clicker questions is very similar to the corresponding exam question (d). However, in order to be able to answer correctly, the student cannot base the exam answer on just remembering the answer from the clicker test; he or she needs to understand the algorithm. In the second example there are two clicker questions corresponding to the single exam question (c). In this case the algorithm involved is much more complex and the student needs also, in this case, to have a thorough understanding of it in order to be able to answer the exam question correctly.

Clicker question:
A program uses 256 KB of internal memory. The page size is 4 KB. How many pages are there in the page table?

1. 64
2. 128
3. 256
4. 512
5. 1024

Exam question:
(d) A program uses 512 KB of internal memory. The page size is 4 KB. How many pages are there in the page table? Explain briefly.

Exam answer: There are 512 K / 4K = 128 pages.

Answer number 1 of the above clicker answers is the correct one and a student answering exactly the same as the given clicker answer at the exam would be given a zero score.

Clicker questions:
A. In which case will the following mutex method for a file not work?

```
while(-f /tmp/lock) {}x
'touch /tmp/lock';
# Writing to a common file
'rm /tmp/lock';
```

1. If a context switch occurs immediately before the lock file is made and another process runs the same code.
2. If a context switch occurs while the other process is writing to the common file and another process runs the same code.
3. If a context switch occurs immediately before the lock file is removed and another process runs the same code.
4. If a context switch occurs immediately after the lock file is made and another process runs the same code.
5. If a context switch occurs immediately after the lock file is removed and another process runs the same code.
6. It will never work.

B. In which case will the following mutex method for a file not work?

```
'touch /tmp/lock';
while(-f /tmp/lock) {}
# Writing to a common file
'rm /tmp/lock';
```

1. If a context switch occurs immediately before the lock file is made and another process runs the same code.
2. If a context switch occurs while the other process is writing to the common file and another process runs the same code.
3. If a context switch occurs immediately before the lock file is removed and another process runs the same code.
4. If a context switch occurs immediately after the lock file is made and another process runs the same code.
5. If a context switch occurs immediately after the lock file is removed and another process runs the same code.
6. It will never work.

Exam question:
(c) Assume that a Perl program, which may be run simultaneously by several independent users, uses the following method in order to avoid two users writing to the same file at the same time:

```
rm /tmp/lockfile ; # Removing /tmp/lockfile
while(-f /tmp/lockfile) {}
# Writing to a common file
touch /tmp/lockfile ; # making /tmp/lockfile
```

In which case will this method not work? Explain briefly.

Exam answer: The method will never work. It removes the lock, checks that it is gone and then enters the critical section. An outright intrusion. Afterwards a new lock is made.

For clicker question A, answer number 1 of the above clicker answers is the correct one and for clicker question B, answer number 6 is the correct one. A student answering exactly the same way as any of the given clicker answers at the exam would be given a zero score as an explanation is *indispensable* when answering 'It will never work'. In the case of clicker question B, the algorithm will never work because of a deadlock situation, in the exam question the algorithm never works for a completely different reason as explained above.

*iii.* Example questions from the category *not related*

During the lectures where clicker questions related to the same overall subject, but where no clicker questions were directly related to the following example question of the category *not related*.

Exam question: A CPU executes an x86 instruction, which adds the numbers contained in two registers. Will the MMU be involved when this instruction is executed? Explain briefly.

Exam answer: No, the instruction has already been fetched and since two registers are added there will be no need to fetch anything from the internal memory and the MMU will not be involved.

**Reidar Kvadsheim** is a retired Associate Professor at the Oslo and Akershus University College of Applied Sciences, Norway. He received his Mag. Art. degree in sociology and Dr. Philos. degree in psychology from the University of Oslo. His research interests include, among other things, behavioral choice, learning processes, and the design of effective and efficient learning support systems.

**Hårek Haugerud** is an Associate Professor and currently Program Coordinator for the Master's Program in Network and System Administration at the Department of Computer Science at the Oslo and Akershus University College of Applied Sciences. He received an M.Sc. and a Ph.D. in Theoretical Physics from the Department of Physics at the University of Oslo. He has been teaching Operating Systems at the Department for a decade and his research interests include virtualization, cloud computing and computer security.

**Hugo L. Hammer** is an Associate Professor at Department of Computer Science at the Oslo and Akershus University College of Applied Sciences. He received a Ph.D. degree in statistics in 2008 and an MEngr degree in industrial mathematics in 2003, both from the Norwegian University of Science and Technology. He teaches engineering students in mathematics and statistics.

**Alfred Bratterud** is an Assistant Professor at the Department of Computer Science at the Oslo and Akershus University College of Applied Science and a Ph.D. candidate within the field of Cloud Computing at the University of Oslo. He also holds a B.Sc. and an M.Sc. in Computer Science from the University of Oslo. He has 10 years industrial experience as a software engineer using various languages, and teaches C++ programming and cloud computing related subjects, at bachelor and master level.

**Laurence Habib** is a Professor in Information Technology and Learning and currently Chair of the Department of Computer Science at the Oslo and Akershus University College of Applied Sciences. She received her M.Sc. and a Ph.D. from the Information Systems Department at the London School of Economics. Her research interests include technology-assisted learning, curriculum development and the universal design of information technologies.