The use of electronic voting systems to enhance deep learning

D. Barreto
School of Engineering and the Built Environment, Edinburgh Napier University, UK

ABSTRACT: The present paper introduces a technology-enhanced teaching method that promotes deep learning. Four stages that correspond to four different student cohorts were used for its development and to analyse its effectiveness. The effectiveness of the method has been assessed in terms of examination results as well as results obtained from class response system software statistics. The evidence gathered indicates that the method developed is very effective and its implementation is straightforward. Furthermore, its success in achieving results seems to be independent of the skills and/or experience of the lecturer.

1 INTRODUCTION

Hand-held devices called Electronic Voting Systems (EVS), Class Response Systems (CRS) or simply clickers have become very popular in recent years and are used for teaching in almost every discipline. In their simplest form they are credit card sized instruments with buttons that students can press to select an option or answer from a multiple choice question that has been previously shown in the form of a presentation slide. Each clicker communicates via radio frequency (RF) or infrared (IR) to a receiver connected to the computer. Application specific software then processes the received signal and it is possible to provide statistics related to the response of the students. Commonly a bar chart indicating the percentage of responses corresponding to each of the possible answers is produced immediately after polling is finished. Students then have a form of immediate feedback that easily demonstrates where they are in relation to the rest of the class.

Recent advances in communication technology mean that transmission of alpha-numerical signals is already possible, even via mobile phones. However, to the knowledge of the author this option is not widely used in higher education due to cost and implementation issues. The discussion presented in this paper is therefore limited to the simplest and most economical version of CRS which allows selection from a list of answers of a multi-choice question prepared in advance.

Unlike an old-fashioned multiple-choice question exam, using CRS provides immediate feedback to students and as stated in various publications they have proven to encourage student engagement (e.g. Judson & Sawada 2002; Hall et al. 2005; Fies & Marshall 2006). Furthermore, positive student feedback in courses where CRS are used is commonly highlighted (e.g. Stowell & Nelson 2007, Kay & LeSage 2009). However, the fundamental deficiencies of a multiple-choice question exam are still present in any (basic) version of CRS. Namely, it is difficult to differentiate when a student has given a correct answer because he/she understood the question or if the answer chosen was guessed or randomly selected. The main problem in the use of CRS in higher education is then related to ensuring that they enhance deep learning (i.e. understanding) while discouraging surface learning approaches (i.e. memorising, fact acquisition, “bottle filling”, etc).

Publications on CRS are abundant. Beatty and Gerace (2009) stated that existing literature tended to fall into three general (and often overlapping) categories: (i) introductions to technology; (ii) reports of individual efforts to teach with clickers and (iii) compilations of recommendations. They also stated that the discussion regarding the pedagogical use of EVS was limited and that how response systems can promote deep learning was an issue that was neglected in the existing published studies. This statement is still valid.

This paper starts by briefly discussing some of the existing CRS-based approaches and how they can be evaluated. The description of the new method, which is based on ideas from the existing approaches, is then followed by details of its development. Subsequently, an evaluation of the effectiveness of the proposed method is presented. This evaluation depends on final examination results, student feedback and CRS software output. The paper concludes with some discussion and conclusion remarks.

2 CRS-BASED APPROACHES AND THEIR EVALUATION

2.1 Evaluation of teaching methods using CRS

Existing research is in agreement that the success of any pedagogical approach when using CRS should be
demonstrated by its effectiveness to foster deep learning (e.g. Roschelle et al. 2004). However it is also important to consider the ease of implementation.

In relation to the effectiveness of a pedagogical approach that uses class response systems, Fies & Marshall (2006) highlighted that missing from CRS research reports are tightly controlled comparisons in which the only difference is the use, or lack of use, of a CRS. The method of application for the use of CRS presented here follows such a research approach. Only one of the variables that are likely to affect the outcomes of this study was changed for each cohort. Hence, under ideal conditions and ignoring environmental factors and differences in the time of lecture delivery, it can be said that the effect of each of the variables can be isolated. Note however that is difficult to verify that all variables are independent of each other. As discussed later, even with these assumptions the addition of extra variables in the analysis seems to affect the performance/effect of the other ones.

An issue that seems to be neglected in most published studies is the ease of implementation of the proposed approaches for CRS use. As a minimum, three questions should be considered: (i) Can the approach be easily implemented by a person with limited teaching experience or lack of pedagogical knowledge? (ii) How is the time required for lecture preparation affected by the proposed approach? And (iii) How is lecture delivery affected by the method (also referring to class size and the need to distribute and collect handsets, or not, at the start and end of the lecture)? With regard to question (i) above, it should be emphasized that lack of experience or pedagogical knowledge are not desirable attributes for any teaching method. It is however advantageous to create teaching methods that can be used successfully by people even without such attributes. Clearly, the availability of experienced and knowledgeable staff can only add to the success of any method.

For the first time, the evaluation of the methods presented here assesses the effectiveness of various methods considering all these factors. Similarly, the new method proposed in Section 3 will also be assessed with reference to these different aspects.

2.2 Existing approaches for CRS-based teaching

Beatty & Gerace (2009) stated that they were aware of only three separate efforts to present and justify an explicit, coherent pedagogy for CRS-based teaching. These were: (i) Peer Instruction, (ii) Question-Driven Instruction and (iii) Technology-enhanced formative assessment. These methods are briefly described here because the new method described in section 3 takes some elements and ideas from them.

A CRS-based teaching approach not discussed by Beatty & Gerace (2009) that needs to be mentioned before those highlighted by them is that by Russell (2008). It is important because it is the only method found that specifically tackles the issue of determining when a multiple-choice question answered via a CRS has been completely understood or guessed. He suggested using sets of questions with contradictory paths which ultimately will help to reveal guessed answers. Such approach also emphasizes the potential of using CRS to adapt the contents of the lecture according to students’ responses, an advantage that is also highlighted in other studies (e.g. Beatty & Gerace. 2009, Dufresne et al. 2000). Although this is a valid and certainly effective approach that fulfills that objective, it clearly requires excessive time for lecture preparation as each answer needs to be treated separately, deriving into numerous complicated alternative paths. Hence practical implementation is difficult, especially in those situations where the use of CRS-based teaching is intensive.

2.2.1 Peer instruction

This approach was proposed by Mazur (1997). The method suggests the use of multiple-choice conceptual questions at strategic junctures during the lecture. When a question is answered incorrectly by a high percentage of students, the class is asked to discuss the question amongst them and then answer again. Quantitative evidence, primarily from pre/post testing (Hestenes et al. 1992) supports that the method improves student understanding (i.e. fosters deep learning). In terms of implementation, the approach seems straightforward and it is widely used at many higher education institutions, but requires lecturers with adequate skills and experience to effectively engage students in discussion. Note however that if a question is answered correctly by a high number of students, there is no reflection regarding the possibility of multiple guessed answers. Furthermore, a correct answer also excludes the need for discussion and peer instruction.

2.2.2 Question-driven instruction

This method, also referred to as the Assessing to Learn (A2L) approach was proposed by Dufresne et al. (2000). In this approach, a “question cycle” or iterative pattern is proposed. In that way the students read a question, think about it alone and/or discuss it in small groups, enter responses, then view the chart of response counts, present and discuss arguments for various choices, and then listen to an appropriate “closure” to the cycle.

An important difference between Peer Instruction and A2L is that Mazur’s approach is intended for intermittent insertion within more traditional teaching methods. The A2L method is intended as the basic structure of class activity, other traditional teaching methods are only used when needed and motivated by the questions and discussion.

Leonard et al. (2001) demonstrated the effectiveness of A2L to foster deep learning but its implementation is not easy. It is not a traditional way of teaching and requires significant experience. Hence some of the evaluating parameters discussed in section 2.1 are not satisfied.
2.2.3 Technology-enhanced formative assessment

This method, also denoted as TEFA is based on the A2L approach and was proposed by Beatty & Gerace (2009). TEFA specifies an iterative cycle of question posing, student discussion prior to selection of answers, post-discussion based on the responses without revealing the correct one and finally, a summary, micro-lecture or closure is provided including meta-level communication. The content of the final closure is normally determined by the previous part of the cycle.

The method differs from others in various ways: (i) teaching is question-driven as for A2L but demands for questions to be challenging, multifaceted and disputable (i.e. no questions in the “you know it or not” style), (ii) opposite to Mazur’s approach peer instruction is encouraged in all questions before and after the students have answered, and (iii) meta-level communication suggests a deviation from the question itself while focusing discussion about learning the content, commenting on the purpose, design, and unfolding of the course itself.

This method has been under development for almost 15 years. Beatty & Gerace (2009) state that it is consistent with established thinking in educational research and that it has also proved to be effective promoting deep learning. Note however, that its implementation is not easy. The creators of the method state that professional development programs are constantly focused with the aim of helping teachers to master the approach. Furthermore, in an engineering context ignoring the development of “you know it or not” style questions is not straightforward. The nature of engineering subjects implies that if something is known the product will be successful and the structure will be safe, if something is not known failure and often catastrophic consequences could occur.

3 A NEW CRS-BASED TEACHING APPROACH FOR ENGINEERING SUBJECTS

The method proposed here has taken ideas from those approaches described in section 2. The method cannot be defended on the strength of experimental findings. Although the empirical evidence presented in sections 4.3 and 4.4 seems to demonstrate that it is successful, the amount of evidence is still limited. The method is presented for consideration as an easily reproducible and implementable method that hopefully can be used by most people with limited experience. It is hoped that the method is consistent with the perspectives of other researchers and users of CRS-based teaching. As in the case of TEFA, the method is based on established ideas in educational research.

The method, as in the case on Mazur’s Peer Instruction, uses questions at strategic junctures of more traditional teaching methods. Hence, the difficulties associated with Question-Driven Instruction present in the A2L and TEFA approaches are avoided. The method however, takes important elements of these approaches and proposes a simple strategy to determine if a question has been understood or guessed, which in contrast to the approach by Russell (2008), has a straightforward implementation. Finally, meta-level communication as discussed by Beatty & Gerace (2009) is also used.

Some researchers suggest that CRS-based teaching is more effective when used as a small part of a lecture when the learning of key concepts is required (i.e. Draper & Brown 2004). The use of CRS should be limited to a single set limited number of questions, especially when experience on their use is limited. This is a suggestion but the success of the method is not expected to be dependent on such restrictions. It has been found from experience (details provided in section 4) that for each set of 5–7 questions the following steps should be carried out:

1. Use meta-level communication to introduce the purpose of the exercise, to explain what can and should be gained from it, and how the concepts learnt can be used when knowledge is finally obtained.
2. Present the question with the possible answers and poll the answers only after peer instruction amongst the students has occurred.
3. Reveal the right answer and show the statistics provided by the software.
4. Explicitly explain how each of the answers was obtained highlighting the mistakes in those that are not correct.
5. Provide a closure, clarification or meta-level comments if required and allow the students to ask more questions.
6. Proceed to the next question and follow steps 2 to 5.
7. Create a question that asks the students how many questions they answered correctly and show the statistics to the group.
8. Create a question that asks the students how many questions they guessed and show the statistics to the group.
9. Assess the outcome of the last two questions and determine how the lecture should proceed.

The use of meta-level communication in step 1 is intended to get the benefits of the TEFA approach. Therefore, as stated by Beatty & Gerace (2009), it should help the students to develop meta-cognitive skills and should help them in the learning process. Note, however, that for the new method proposed here this is done at the start of the questions and not at the end as in the TEFA approach.

Peer instruction in step 2 reaps the benefits to foster deep learning as suggested by Mazur (1997). It should motivate and engage the students with the subject.

Step 3 provides the usual CRS feedback for students and allows the students to identify their position in relation to the class. Note however that this only gives a measure of performance in a single question.

Step 4 is a very important one and generally allows emphasizing the causes and consequences of failure. This is very useful in engineering contexts. At this
stage meta-level communication in addition to that presented in the following numeral is possible.

The two final questions referring to the number of correct answers and the number of questions guessed are very dependent on the honesty of the students, but they are very consistent and provide an easy way to verify if the whole exercise was productive, effective and if the topic was understood or if positive results were the results of multiple guessed answers. Furthermore, simple statistical analyses after the lecture allow determining if what the students stated in these questions coincides with their answers to previous questions. Honesty in the students’ answers would reflect a match between the number of correct answers and the percentage of correct answers obtained in the previous sections. Although data regarding this comparison is not presented in this paper, it suggests that the student cohorts assessed in this study were honest. These two steps also show the students a much more complete picture of their understanding with respect to the rest of the class in comparison to what a bar chart from the results of a single question can offer them. Ultimately, the analyses provide enough information to determine how the following lecture should be approached and ensure that the fundamental concepts required are clear if necessary.

When formulating the new method a rigorous research approach was followed. In particular, the comments by Fies & Marshall (2006) regarding tight control on the evaluation of the effectiveness were considered. The final method as presented above is the result of using ideas by others researchers in an incremental fashion. Four stages of development were assessed, which correspond to experiences with four different student cohorts. The details of each of the development stages and the measures used to evaluate the success of the method are presented in the next section.

4 DEVELOPMENT AND EVALUATION OF THE METHOD

Four groups of students were used to develop the method. All groups were 3rd year undergraduate Civil Engineering students as part of a module in Geotechnical Engineering. The effectiveness of the method was evaluated in relation to the teaching and learning of bearing capacity.

In simple terms bearing capacity is the ability of the ground to support the loads transmitted to it from the structures built on it. Being able to calculate the bearing capacity of a certain soil under varying loading conditions is a key learning outcome in any geotechnical course. The nature of the subject and the existence of different geotechnical design codes worldwide require adequate knowledge of various definitions (types) of bearing pressure/capacity. While standard calculations of bearing capacity are given in gross terms, considering situations where the foundation depth is great it might be important to consider the value of bearing capacity in gross effective or net effective terms. The difference between these definitions are directly related to the definition of Terzaghi’s principle of effective stress, a corner stone of soil mechanics which is also recognised to be a difficult concept to teach. Similarly, it is of the utmost importance to understand the difference between an allowable bearing capacity, an ultimate bearing capacity and a presumed bearing capacity, especially when referring to accepted values quoted in design codes. The corresponding definitions for these terms are not presented here but they can be found in most textbooks on soil mechanics and foundation engineering. The omission of such definitions should help to emphasize that the teaching method presented in this paper is equally valid for any topic, but it is expected that is particularly effective for engineering subjects.

4.1 Details of module delivery

Bearing capacity is only one of the topics in the module. For the four student groups all the topics of the module were delivered to them using a traditional teaching approach involving a combination of 12 two-hour lectures, 6 tutorial sessions including formative assessment, 3 laboratory sessions counting for 10% of the module marks, a piece of coursework counting for 20% of the module and a final exam providing the remaining 70% of the marks.

The first student group was taught with no use of CRS, while the other three groups included a lecture where a different CRS-based teaching approach was used to teach the different definitions of bearing capacity/pressure. Further details of the different methods used are provided in section 4.3. Although the number of students and time of delivery was different for each of the cohorts, the course contents, mode and pace of delivery remained unaffected. Evidence presented in section 4.4 as well as student feedback suggests that these differences did not affect the outcomes of this research.

4.2 Method evaluation measures

In section 2.1 it was indicated that CRS-based teaching should be evaluated in terms of its effectiveness to foster deep learning and also the ease of its implementation. The details provided in section 3 clearly demonstrate that the method is implemented easily. The use of CRS-based teaching is only present on a small section of the whole module delivery. The type of question needed in the method does not differ from that normally provided in any tutorial problem or exam question. Furthermore, the preparation of the CRS-based questions requires only little extra preparation in contrast to that usually required with other approaches such as those by Russell (2008), Dufresne et al. (2004) and Beatty & Gerace (2009), where the use of CRS is very intensive. Hence, significant staff experience in the use of the clickers for teaching is not required. The only difficulty that may arise in terms of implementation will occur if the number of students is large.
Note, however, that this is a logistical problem but it does not seem to affect the effectiveness of the method as highlighted in section 4.4.

In terms of effectiveness to foster deep learning, examination results and statistics provided by the CRS software were the main sources of information used to evaluate the benefits of the method. The final exam for the four student cohorts consisted of five questions. One of them was related to the topic of bearing capacity. The question consisted of a mix of calculation and theoretical/conceptual sections to ensure verification of understanding and to facilitate the detection of surface approaches to learning (i.e. memorised responses or mechanistic procedures to find an answer without considering its general context). Comments by various internal/external examiners during the exam moderation process have highlighted the good quality of the question and how it is industrially relevant and useful to evaluate the achievement of the learning outcome specified. The question is therefore transcribed below based on such comments and considering the main audience for this paper. Note, however, that the method proposed in this paper could be successfully used for any topic and discipline.

Exam question: “The bearing capacity of a pad foundation can be calculated using the formula given by:

\[ q_f = cN_c s_i c + \gamma DN_q s_i q + 0.5 \gamma BN \gamma s_i \gamma \]

The derivation of this equation is based on the occurrence of a certain failure mechanism (i.e. general shear failure).

(a) Make a diagram of the general shear failure mechanism and refer to it to explain the meaning of the three terms in the equation above.

(b) “For geotechnical engineers the bearing capacity when a local shear or punching failure mechanism is likely to occur is not their biggest concern”. Comment if you agree or disagree with this statement providing reasons for your answer.

(c) Ignoring depth factors determine the drained gross bearing capacity of a 2.5 m × 3.5 m pad foundation placed at a 2 m depth. The soil found in situ is a firm, saturated clay (\( \gamma_{sat} = 21.5 \text{kN/m}^3 \)) layer extending to considerable depth with the following geotechnical properties: \( c' = 3 \text{kPa} \) and \( \phi = 27^\circ \).

(d) Calculate the drained bearing capacity of a foundation as described in part (c) but considering (i) gross effective, (ii) net and (iii) net effective terms.”

The nature of the question implies that plenty of formulae are required in order to calculate bearing capacity, shape and inclination factors. It was of course intended to avoid the encouragement of mechanistic procedures and encourage deep learning, so these formulae were provided on an additional sheet. It is worth mentioning, however, that definitions and formulae related with part (d) were obviously not included on this sheet as they correspond to fundamental principles that need to be conceptually understood and therefore evaluated. Furthermore, they are the basis of the CRS questions detailed in Section 4.3.2 which are also used to evaluate the effectiveness of the proposed methods.

In the UK, each cohort is usually provided with past exam papers of previous cohorts for study and revision previous to the exam. This was also done for the four groups analysed here. It is worth emphasising here that although the four cohorts did not answer the exam simultaneously as explained in Section 4.1, this did not mean (for example) that the 4th cohort had knowledge of the exams given to the three previous cohorts. This is an obvious measure that has to be taken to maintain the objectivity of the research approach. Since all the cohorts needed to be evaluated using the same exam questions, precautionary measures were taken to ensure that the students did not memorise the procedures or answers (i.e. an additional set of different review questions that asked the same subjects were provided). Additionally, in the final examination, the question on bearing capacity was identical for the four groups and only differed on the numerical values (i.e. parts c and d) and slight deviations in theoretical questions were also introduced where this was possible (i.e. parts a and b) without changing the learning outcomes that were assessed. Plagiarism was rigorously prevented and never detected. Furthermore, as highlighted above, none of the cohorts knew at any stage that their results were being compared to those of other cohorts and of course, they never had access to the exam questions of the previous cohorts. So the first measure of effectiveness of the method was the results (marks) obtained by the students on the question about bearing capacity.

It is also common practice to allow the students to choose 4 of the 5 questions to be solved. This approach was also taken for the four student groups. As a consequence, it is expected that students will choose to solve the questions they feel more confident with. That is because students are expected to intuitively answer those questions which they believe will help them to get the highest marks. Hence the percentage of attempts for the question can also be considered as a measure of effectiveness of the method described in section 3.

In relation to the CRS software statistics, the percentage of correct answers is a very intuitive and obvious manner to assess effectiveness. However, in this study, the responses provided by the students to the questions related to the number of correct answers and the number of guessed questions were also considered. It is believed that an increase in the number of correct answers accompanied by a decrease in the number of guessed answers is demonstration of improvement and effectiveness of the corresponding method.

4.3 Development stages

The development process involved four cohorts of students. Each of the cohorts corresponds to a certain development stage which in turn includes a particular enhancement which finally derived in the final method.
4.3.1 Stage 1 – The control group (No CRS)

In order to prove if the use of clickers is effective or not it is necessary to compare the results of a student group taught without using the system with another cohort using CRS under otherwise identical conditions. The control group for this stage (with no CRS-based teaching) consisted of 24 students.

For this group, each of the different definitions of bearing capacity to be taught were introduced to the students as part of a traditional lecture session. Each definition was defined appropriately in a single presentation slide followed by one or two more slides detailing a solved calculation example illustrating the principles behind each definition. After each example was presented the students were allowed to ask any questions or express any doubts or concerns. However, if there were no questions the lecture proceeded as normal. This is an approach very similar to that suggested by Mazur (1997), but with the absence of CRS. Since at this stage rapport between the lecturer and the students had already been established it was concluded that the concepts had been fully understood by the students if no doubts/concerns were raised after the concepts were introduced (as it happened in most of the definitions taught). Surprisingly, the examination results revealed the opposite.

Since CRS-based teaching was not used, the only comparable measures for the evaluation of effectiveness are related to students’ results in the final examination. Results indicated that 66.7% of the students attempted the question on bearing capacity, making it the least attempted question. This compares to 91.7% of attempts for the most popular question. Those students who attempted to answer the question obtained an average mark of 38.0% with a standard deviation of 24.0%. Clearly, although the range of variation of results was significant, the achievement of learning outcomes is in general terms not satisfactory.

The lack of questions during lecture time and the exam results imply that although the students are able to follow and understand the principles explained as shown by the presented examples, they are not able to apply or extrapolate such concepts to different problems. Hence, using such an approach, the fostering of effective deep learning cannot be guaranteed. As a consequence, the use of alternative teaching methods as part of the more traditional lecture is justified.

4.3.2 Stage 2 – The introduction of CRS

The second student group consisted of 80 students and was taught under the same conditions of that at Stage 1. The only difference between the two cohorts was the use of CRS-based teaching in the lecture dedicated to the different definitions of bearing capacity. This meant that following the slides describing each definition, the calculation example was shown. Then, after time for questions and comments was given, a slide with a multiple choice question/problem to assess understanding was introduced and the clickers were used by the students. The problem asked was the following:

“The vertical load on a 4 m wide strip footing (including its self-weight) built on sand will be 400 kN/m². The footing base is 2.5 m below ground level and there is a static ground water table 1 m below ground level. The sand has a unit weight of 19 kN/m³ when saturated and 16 kN/m³ when dry. Calculate (a) the gross bearing pressure (q_{gross}), (b) the overburden pressure (p_0), (c) the net bearing pressure (q_{net}), (d) the gross effective bearing pressure (q'_{gross}) and (e) the net effective bearing pressure (q'_{net}).”

Hence, parts (a) to (e) in the problem above corresponded to a single multiple choice question to be solved using the clickers by the students. Solutions to these answers are not included here and neither are the possible answers for each of the questions. However, amongst the answers in the slide corresponding to each of these questions there was a correct one and various incorrect values that could be found if conceptual mistakes or common arithmetic errors were committed. The students were only asked for the correct answer and an appropriate (measured) amount of time for the solution of each question was provided. After the answers were polled the immediate feedback as provided by the CRS software showing the percentage of answers for each option was shown to them. Subsequently a few seconds were given for them to digest their own position in comparison to the rest of the class. Student feedback indicated that this stage was extremely valuable for the students to realise if they had actually understood the concept or not. So, it is clear that actually demanding the application of a concept in a different scenario to that used for its explanation is useful. Although the CRS are not an essential requirement to do this, they serve this purpose very well as they stimulate student engagement as demonstrated by various researchers (e.g. Draper & Brown 2004, Hall et al. 2005). However, as discussed before, this approach is not effective to differentiate which questions were understood completely or guessed even when the percentages of right answers are high. Hence, the simple approach described in steps 7–8 of section 3 was also introduced at this stage of the lecture. This could also be done independently of the CRS, but their use makes this process much easier, and makes the information available for further analyses after the lecture.

The exam results for this stage were very interesting. The percentage of attempts rose from 66.7% in Stage 1 to 80.0%. Similarly the mean value of the marks obtained also increased from 38.0% to 47.2% while the standard deviation remained almost constant at 23.6% (compared to 24% at Stage 1). Results were therefore encouraging; they indicated better student performance and higher student confidence. Student feedback related to the use of CRS was also positive. Furthermore, the constant standard deviation and the
range of results seemed to indicate that the preventative measures taken in relation to past exam papers had been effective. Most importantly, all the results implied that the introduction of CRS-based teaching and the pedagogical principles described above had resulted in improved students’ learning.

4.3.3 Stage 3 – CRS and peer discussion
The third group of students was comprised of 37 students. Following the same approach, teaching methods were exactly the same to those during Stage 2, and differed only on the introduction of peer discussion as suggested by Mazur (1997). In practical terms this meant that a period of time was introduced before polling of answers was accepted. During this time, students were asked to discuss amongst them about the answers and to argue and defend their positions within small groups before any answers were accepted by the CRS software. Hence, at least in theory, any changes in examination results could be attributed to the introduction of peer instruction only.

Interestingly, examinations results indicated a slight decrease in the percentage of attempts from 80.0% to 75.7%. Note however that such percentage is still high in comparison to that found during Stage 1 (66.7%). Similar observations could be made in terms of the marks obtained. The mean value was 46% (a slight reduction from 47.2%). On the other hand, the standard deviation was 24.4% (very similar to that in both Stage 1 and 2). Note that the number of students during Stage 2 was 80 compared to 37 during this stage. This seems to indicate that the results are not affected by changes in the size of the student group as discussed before. Also, it would not make sense to conclude, that in light of the results obtained, the introduction of peer instruction had no effect on the effectiveness of the method as this would contradict a significant body of evidence available (e.g. Hestenes et al. 1992, Mazur 1997). However, it can be said that peer instruction is perhaps the process that has the highest dependence on the experience of the lecturer. This also requires skill and although it can be confirmed that there was discussion amongst the students, perhaps more control and guidance was required from the lecturer at this stage to ensure its’ effectiveness.

4.3.4 Stage 4 – Final conception of the method
The cohort at this stage consisted of 24 students. As before, the method of delivery remained identical with exception of a unique variable. In this case the variable of study was the effect caused by the addition of meta-level discussion. As described in section 3 this was made at the very beginning of the CRS-based teaching method instead of the approach of Beatty & Gerace (2009) that does so at the end of a question cycle. This was considered to be a risky decision because it was thought that telling the students that they were going to be asked, and how they were going to be assessed would put them behind the achievement of a short-term and perhaps forgettable outcome once it was finished. In the next section evidence will illustrate that perhaps there might be evidence of such phenomenon because the statistics provided by the CRS software show significantly higher results in Stage 4 when compared to those than in the earlier stages. However, the examination results are surprising, very encouraging and suggest that the final structure and methodology of the CRS-based teaching proposed is very effective.

The percentage of attempts rose to 100%, making the bearing capacity question the most popular one in the exam for the first time. Similarly, the mean value of the results increased to 72.5% (from around 46% in Stages 2 and 3) with a much smaller standard deviation of 8.3% (in comparison to approximately 24% in the previous stages). The results are not only significantly higher, but they also demonstrate that the range of results is smaller, confirming that the learning outcome had been achieved by every student.

In spite of the results, there is not enough evidence to conclude that the success of the method is due to the introduction of meta-level discussion, even if this was the only variable that changed at this stage. The reason for that is thought to be that all variables might be influenced by each other. For example, meta-level discussion can give students a purpose, and in turn this will aid or encourage peer discussion. Hence, the results do indicate that the combination of methods in the final approach described in section 3 is very effective and does promote deep learning. As an added bonus, the teaching approach is very easy to implement. Hence, although more evidence is still required, the method is introduced as an easily implemented alternative with the likelihood of a good outcome, which also seems to be independent of the lecturer’s skills and experience.

4.4 Evaluation of CRS software statistics
All the conclusions and statements provided so far were based on the results of final examination marks. It was of particular interest to analyse if the same conclusions could be derived when the CRS software statistics were studied. Figure 1 shows the percentages obtained by the students for the correct answer to 5 questions at the three stages when the corresponding

Figure 1. Percentage of students selecting the correct answer for five different questions at various stages of method development where CRS-based teaching was used.

189
Figure 2. Analysis of verification questions to differentiate surface and deep approaches to learning.

CRS-based teaching methods were used. It can be clearly seen, that excluding questions 1 and 2, the tendencies and conclusions derived from examination results are confirmed. That is, Stage 2 is the least effective, followed by subsequent increases on effectiveness for Stages 3 and 4 with Stage 4 being the most effective.

It is also interesting to see the results obtained from the verification questions (i.e. those to determine number of correct answers and number of questions guessed). Figure 2 illustrates the evolution of these parameters for each of the stages when CRS-based teaching was used. Note, however, that the figure illustrates the percentage of students that stated answering all questions correctly and the percentage of students confessing that they had to guess all the questions’ answers. It can be seen that there is a slight decrease in the percentage corresponding to guessed answers as the method becomes more sophisticated (i.e. as the Stage No. increases). Similarly, there seems to be an overall increase in the percentage of students getting all the answers correct. This is relevant because examination results presented in the previous section seemed to suggest that there was no sign of improvement from Stage 2 to Stage 3 when peer instruction was introduced into the method. Figure 2 does confirm that peer instruction did have a positive effect demonstrated by the reduction in guessed questions, even though there is a slight reduction in the percentage of students getting all correct answers from stages 2 to 3.

As discussed in Section 4.3.4, the results in Figure 2, and in particular the radical increase in the number of correct answers from stage 3 to 4 might pose the question of the validity of deep learning. Such a change is of course a positive outcome for the method, but with the evidence available it is very difficult, if not impossible to determine whether these results are affected or not by surface learning approaches. Note, however, that exam results seem to support the conclusion that deep learning has been achieved, and hence that the method is successful.

In contrast to Figure 2, a continuous increase (for all stages) in the percentage of answers answered correctly is observed when the percentages are assessed in an aggregated manner. That is when the percentage related to getting some (but not all) of the answers correctly (or guessed) is also considered. Figure 3 shows the results considering the percentage of students that answered 80% or more of the questions correctly, together with the percentage of students that guessed 80% or more of the questions asked. When presented in this form, the results clearly show the benefits of the various additions introduced in the method at each of the development stages.

The results in Figures 2 and 3 also indicate that independently of the CRS-based method used, it is difficult to promote student engagement amongst those individuals that are not prepared or that do not intend to engage with the material. Nevertheless, it is encouraging to observe that there was a decrease as the approach became more developed. It is also believed that the results shown in Figures 1 to 3 and the examination results discussed in section 4.3 demonstrate with significant confidence that the new CRS-based teaching method is highly effective to foster deep learning.

5 DISCUSSION AND CONCLUSIONS

The present paper has used an incremental and rigorous research approach to develop and implement a CRS-based teaching method that promotes and enhances deep learning. Four stages that correspond to four different student cohorts were used for its development and to analyse its effectiveness. Each of the stages has attempted to look at effects caused by various approaches widely accepted in the educational literature. Amongst them, peer instruction, question-driven instruction and meta-level communication have been described, discussed and analysed where applicable.
The effectiveness of the method has been evaluated in terms of examination results as well as results obtained from CRS software statistics. It is recognised that the evidence gathered so far might be limited, but it does seem to demonstrate that the method is highly effective.

The main conclusions that can be inferred from the results and evidence presented include:

- The proposed approach is highly effective to promote deep learning, is easy to implement and it is based on educational ideas that are widely practised and strongly supported in the existing literature.
- The approach presented incorporates a very simple method that can be used to determine very easily to what extent the questions have been completely understood or its’ answers guessed.
- There seemed to be a strong inter-dependency between the various approaches/methods that were incrementally added into the method. It was also observed that peer instruction requires a certain amount of experience to achieve conclusive results.
- It was demonstrated that although examination results alone seemed to indicate that there was no effect on the results caused by peer instruction, the CRS statistics demonstrated that such addition is indeed a very relevant part of the newly proposed method.
- It is difficult to verify if the different parts of the method are independent of each other, but it has been shown that as a complete unit, the proposed method works and enhances deep learning.
- Finally, it must be said that at various development stages it is difficult to determine whether the results obtained are influenced by unavoidable surface learning approaches adopted by some students. However, without stating that final examination results fully represent deep learning, it can be concluded that it is most likely that the results reported are the consequence of the encouragement of deep learning approaches. This is because all the necessary steps and measures to ensure that this is done have been considered for this research. These measures were detailed and described in the paper.

REFERENCES


