ELECTRONICALLY ENHANCED CLASSROOM INTERACTION

Stephen Draper
Department of Psychology
University of Glasgow, UK
s.draper@psy.gla.ac.uk

Julie Cargill & Quintin Cutts
Department of Computing Science
University of Glasgow, UK
julie.quintin@dcs.gla.ac.uk

Abstract

A design rationale for introducing electronic equipment for student interaction in lecture theatres is presented, linking the instructional design to theory. The effectiveness of the equipment for learning depends mostly on what pedagogic method is employed: various alternative types are introduced. Prospective studies are outlined for exploring its use over new ranges of application. Rival views of the concept of interactivity are one way to organise the evaluation of this learning technology.

Keywords

creative solutions, overcoming limitations, innovation, convergence, design rationale, niche, feedback, group response

Introduction: The Design

This paper describes the design rationale for introducing electronic equipment for student interaction in lecture theatres, and the studies now in prospect of the use of this equipment.

The equipment is essentially that of the TV show “Who wants to be a millionaire?”: every member of the audience (i.e. each learner in a lecture theatre) has a handset similar to that of a TV remote control, the presenter displays a multiple choice question (MCQ), each learner transmits the digit corresponding to their chosen answer by infrared, a small PC (e.g. a laptop) accumulates the answers, and it displays, via the room’s projection system, a bar chart representing the distribution (totals) of the responses to audience and presenter alike.

This may be called (following Michael McCabe) a “Group Response” (GR) system. Its essential feature is that, regardless of group size, both audience and presenter get to know the distribution of responses (alternatives chosen), and how their own personal response relates to that distribution, but however without knowing who chose what. This means everyone contributes, and the representativeness of each response is also exactly known. On the other hand, the privacy of the choice means that, unlike in face to face groups, each individual can express the choice they incline to, rather than only a choice they feel able to explain and justify to others. These are quite often different both in science learning and in social processes.

The main pedagogic categories of use of the equipment are:

- **Assessment**, both formative and as practice for summative assessment. Here the MCQs are meant to test content knowledge, and perhaps are drawn from a bank used for formal assessment on the course. The advantages of the equipment here are that “marking” is fully automatic, each learner can know immediately if they gave the right or wrong answer, how their performance on the question compares to the group as a whole, tailored explanations may be given by the presenter, and the presenter equally sees immediately how well the class measures up on that question (feedback from learners to teacher). The feedback cycle here takes about two minutes per item (somewhat longer if explanations are given). Any kind of MCQ may be used, provided the response is a single selection from a small fixed set: whether the usual rather shallow item, or one designed to probe understanding more than information retention (possibly by prior use of phenomenography (Marton, 1981; Marton & Booth, 1997) to map the common misconceptions).

- **Formative feedback on learning** within a class (i.e. within a contact period). Similar items might be used, but in order to discover and demonstrate what points should be focussed on during the class. Thus one or several such question items at the start of a class could be used to select a topic for detailed coverage, while the same or similar items at the end could demonstrate to what degree the group now understood the topic.

- **Formative feedback to the teacher** on the teaching (i.e. “course feedback”). While the standard questionnaire at the end of a term, semester, or course has in general only a small effect on changing anything (Cohen, 1980) and takes a year to do so, a quick on the spot anonymous poll half way through a class (e.g. on whether the pace is too fast or too slow, the jokes too numerous or infrequent, the examples too many or few) can be used to change things immediately. Making adjustments to the teaching every 30 minutes, instead of only once a year, and furthermore making them for the particular group that gave the feedback, is much more likely to be effective than the usual practice.

  Even better on the spot evaluation might be done by asking students what the best and worst issues are in the teaching at present. Assuming that even a handful are willing to mention an issue to the teacher’s face, these can then be put as questions to the class, and an accurate secret ballot taken on the breadth of support for each one. This cycle of an open-ended evaluation probe, followed by systematic (and quantitative) measures of the issues thus identified, is the best evaluation practice: much better than using standard course questionnaires for all classes, learners, teachers, and contexts. Normally it would take days or weeks: but the whole 2-phase cycle could be done within 10 minutes.

- **Peer assessment** could be done on the spot, saving the teacher administrative time and giving the learner much more rapid, though public, feedback. For example if each student has to give a verbal presentation and this is peer assessed, then at the end of their talk the teacher can display (say) each of 10 criteria in turn, and get the other students to enter their mark for this anonymously but on the spot, with the totals displayed.

- **Community mutual awareness building**. At the start of any group e.g. a research symposium or the first meeting of a new class, the equipment gives a convenient way to create some mutual awareness of the group as a whole by displaying personal questions and having the distribution of responses displayed. For example, at a research meeting start by asking people’s ages (which illustrates the advantage of anonymity), and the kind of department or institution they come from, and some alternative reasons for attending. At the start of a class, I might ask whether each student is straight from school or not, their gender, which faculty they belong to, whether they signed up for the course because it is their main interest, a side interest, or are just making up the number of courses they do.

- **Experiments using human responses**: for topics that concern human responses, a very considerable range of experiments can be directly demonstrated using the audience as participants. For instance visual illusions may be displayed and the equipment used to show what degree of uniformity of response is found. Priming effects can be shown, where the perception of an ambiguous word or display is affected by what was shown before. The performance of witnesses to a crime (including the effects of some well known biasses) can be explored by showing a short film, followed by various questions about what was shown. Social psychology effects, e.g. on conformity, could be demonstrated if responses to early questions were faked to see whether the class then changed their responses to later questions. In general, experiments that rely only on a stimulus and a forced choice response, but not on accurate measurements of reaction times, can usually be demonstrated in this way. Thus for the particular case of psychology, but also for parts of physiology, medicine, economics, and so on, direct demonstrations of relevant effects can be mounted.

- Possibly the most productive application, however, and the one with the largest body of existing research, is in using the equipment to **initiate a discussion**. Here, a carefully chosen MCQ is displayed and the learners register an answer, thus privately committing to a definite opinion. The presenter then, however, does not indicate the "right" answer but directs the class to discuss their answers with each other. Having to produce explanations and reasons is powerfully "mathemagenic" (conducive to learning), which of course is why researchers learn so much from giving talks and writing papers, and why teachers make their students write essays and answer questions. The equipment can be a significant help in introducing this, even into large classes. This method of teaching by questions has been widely used and researched, although mostly without electronic aids (Hake, 1998a, 1998b).

Justification or Design Rationale

Although techno-enthusiasts, and indeed many government agencies or departments, have been pushing the use of computers and other technologies in education, and there are now many people whose job is essentially this and who are therefore necessarily aligned with this indiscriminately positive attitude, there is still very little good evidence of benefits. Perhaps this is not surprising: Landauer (1995) found it very hard to discover evidence of economic benefits for using computer technology in general. Besides suggesting that developing evaluation methods powerful enough to test this may be a more important, if more difficult, research task than generating yet another application of technology to learning, this does mean that each application should be carefully justified. In a review of a number of applications (Draper, 1998), I argued that most applications showed no significant improvements over what they replaced, but that the few striking positive exceptions were characterised by "niche-based design": by a good fit between a particular learning situation and a specific technical solution. They were projects that had been inspired by identifying a specific weakness in current delivery, and had focussed technology on solving that problem rather than on replacing what had been adequately done before. Can the use of the classroom equipment described above meet the implied standard of justification?

In considering large classes in large lecture theatres, the main problem is usually analysed as to do with the lack of interaction and the consequent extreme passivity imposed on the audience. In terms of Laurillard's model of the learning and teaching process (Laurillard, 1993, p.103), this situation fails to support the iterative interaction between learner and teacher that is one of her underlying principles, and more specifically does not support even activity 2: the "re-expression" by the learner of what the teacher has expressed. (This can be seen as corresponding to the constructivist requirement that learners acquire knowledge by rebuilding it on their own personal, mental foundations. Redescribing it in their own terms is an activity that powerfully promotes this.) Actually, with highly skilled learners and a teacher reasonably in tune with the group, this can nevertheless take place: for instance, where the learners take notes that are not mere dictation, but substantial re-formulations of what is being talked about. (This is a reasonable theoretical analysis of the considerable benefits I have often obtained from listening to talks at conferences where I have not asked questions, but have nevertheless learned something useful.) However this degree of skilled, silent interaction is not often present in undergraduate teaching, and large numbers usually prevent learners asking sufficient questions to repair the attunement between speaker and audience, from both a pragmatic (there isn't time for many people to ask questions) and a social (it just feels too embarrassing) viewpoint.

That, then, is the diagnosis offered here of the chief weakness of lecturing to large groups. The handsets and associated equipment offer a way of tackling that weakness by (a) allowing each learner independently to generate an answer (at least a partial instantiation of activity 2), whereas otherwise only the handful who put their hands up really do this; and (b) to register that answer and so maintain the motivation for doing it; and in so doing (c) to affect the course of what happens next. This contingency (dependence of the teacher's behaviour on what the learners do) is true interactivity: one of the underlying principles of Laurillard's model, represented there by the to and fro repetition of activities between learner and teacher. The summed responses are real feedback to the teacher, that naturally leads to adjustments and reattunement if required, and in fact do this better than questions and answers from any subset of individuals. Furthermore the equipment offers an anonymity of response that addresses the shyness that additionally inhibits any interaction.

As mentioned in passing, there are some other reasons for expecting benefits with the types of pedagogic use other than initiating discussions. Formative, summative, and peer assessment could be made more convenient and quicker (and so more affordable for both learners and teachers in terms of time). Starting to build a sense of a learning community could get off to a quicker start, especially in large groups. Demonstrating experimental effects instantly connects the abstract overview given to a personal perception and experience of it: something very helpful to learning both for retention, comprehension, and for a fuller content of learning. The biggest learning gains, however, are likely to come from the much better and quicker feedback from learners to teachers, allowing better attunement of the delivery; and from the method of teaching by questions i.e. of discussions in class (whether in small groups, plenaries, or a combination) initiated by well designed questions and by getting each individual to start by committing to an initial position.

Is the equipment really likely to be any better than the alternatives? The simplest alternative is getting students to give a show of hands. This equipment crucially offers more privacy (it's a secret ballot, and important for just the same reasons). Other rival technologies are to issue each student with a cardboard or plastic cube with a different colour on each face, to be turned to show their "vote"; or with a large sheet of paper divided into a few squares each with a digit in, that the student can hold up in front of their bodies and point to the digit they select. These methods allow only near neighbours to see a student's selection. Thus the electronic equipment offers somewhat better privacy, but the difference may only be crucial with new classes: it is quite possible that with a class grown comfortable with the electronic version, moving over to a cheaper but less private version might not destroy the interactivity. The electronic version also provides faster and more accurate counting of the results: most presenters will only estimate shows of hands to about the nearest 20%, unless they have the patience to pursue and count exactly even with large groups. The accuracy may have a small but not negligible value in making all participants feel their views count, and are not just lost in crudely approximate estimates.

In scrutinising this instructional design rationale, note that it does not feature computers in a starring role (although actually one is crucial to tabulate the results): the instructional design mostly isn't in the equipment or software, but in how each teacher uses it. That is a lesson which perhaps the rest of the learning technology field should take more to heart if the aim is in fact to improve learning rather than to promote the glamour of machines. On the other hand, note too that this design does not fit with a simplistic interpretation of the slogan "learner-centred". Improved learning and the learners are the ultimate intended beneficiaries, but one of the important ways that end is achieved is by first serving the teachers better, by giving them much better, faster, and more detailed information on what the learners are thinking now, and where their problems are at each point.

Prospective Explorations

There is a considerable history and community of practice in using such equipment in the specific area of promoting discussion (the last of the pedagogic uses listed above) and so improving student understanding in science and engineering at the school and early university levels (e.g. Hake 1998a, 1998b). The authors have obtained sufficient equipment for several lecture theatres, and are about to begin exploratory studies, particularly with a view to exploring the range of applications, and how far its utility can be demonstrated beyond its best established application area. We hope to trial its use in all of the pedagogic modes listed in the first section, in two universities (Glasgow and Strathclyde), in at least two disciplines (psychology and computer science) in both universities together with several others as opportunities arise, at various levels (years) in undergraduate programmes, and in a range of group sizes from 300 students downwards. (The biggest need and the biggest potential gains are in the largest group sizes, but innovation is of course a lot "safer", i.e. easier to manage, in smaller groups.)

The exploratory studies should yield practical knowledge such as question banks for the participating disciplines, and how much support is needed for first time use (a new lecturer and students who haven't used the equipment) and for regular use. They will also yield evaluation results on what benefits can be demonstrated. We hope to use a version of the method of Integrative Evaluation (Draper et al., 1996) to address both these aspects.

Interactivity

According to Jim Boyle (personal communication), students are generally, although not universally, enthusiastic about this approach, even over long periods (e.g. regular use throughout a year). When asked if they regard the interactive equipment as an advantage or not, classes typically show a spread of opinion such as 70% for it, 20% indifferent, 10% definitely opposed to it. Investigating more deeply than general student preferences will require more, and more sophisticated, measures.

Some of the most important evaluation issues can be organised around the notion of interactivity. Some researchers tend to an almost mechanical interpretation of interactivity e.g. counting the number and branching ratio of choice paths for users in multimedia learning software (Sims, 1997; Hoyet, 2000). With this equipment, that corresponds to the number of questions put to the learners for them to respond to, regardless of their content. It also corresponds to the effects we may well see of novelty, of the perception that the teachers are taking special trouble over the teaching (the Hawthorne effect; Mayo, 1933), or simply of physiological arousal (the physical activity involved in pressing buttons i.e. mechanical interactivity) which has led to the heuristic rule of not lecturing for more than 20 minutes without a pause, having the audience move around periodically, etc. On the other hand, if we believe in the Laurillard model, then the important factor would probably be the amount of time each learner spends on activity 2 ("re-expression"): so using the handsets should be better than a non-interactive monologue, but not as good as time spent in peer discussion (open-ended verbal responses rather than selecting one of the digits on the handsets). In other words, the measure of it would be the number of mental and verbal responses a learner makes (in discussion) rather than the number of button presses on the handset. On the other hand again, if what is important about "interactivity" is actually changing what happens by visibly affecting the teacher (i.e. genuine human-human interaction with the actions of one party being contingent on those of the other), then it will be changes to what the session is used for as a result of responses to questions near the start that predict the largest learning gains. Varying approaches in classes, and taking independent measures both of learning and of enjoyment or alertness should eventually allow such questions to be decided. Measures taken over time (e.g. weeks) should allow any halo and Hawthorne effects to be independently identified, if they are present, with enthusiasm decaying as the novelty wears off, or performance being independent of the learning activity tried and only dependent on the perceived interest of the researchers.

Other Technical Details

There are some further detailed issues that arise, and could be investigated. The particular equipment used transmits not only a digit to signal the learner's selected response to the question, but also a confidence level (high, medium, or low), and an ID for that handset which may or may not have been arranged with a known mapping to the student's identity. Furthermore the number of attempts each learner makes at the question before the cut off time may be recorded. The GRUMPS (2001) project is interested in exploring data mining of records of such student interactions, though that involves negotiating issues of privacy and data protection with the students. We are writing software to smooth the integration of the equipment with other lecture facilities (e.g. the use of PowerPoint presentations), and with keeping records of the interactions.

There seem to have been a variety of particular equipment used in the past, and more than one type is currently available. For instance a one-button system has been used (Poulis et al., 1998), though that required each response option for a question to be attended to separately. Various numbers of buttons are offered in other equipment, and sometimes the ability to enter multi-digit responses and transmit them as one number. Wired, radio, and infrared implementations have been used. Currently infrared proves cheapest. Already technically feasible, though not yet financially attractive, is the solution of equipping every student with a radio-linked PDA (e.g. palmtop computer). Functionally, the features that can matter to further pedagogical tactics include: entering multidigit numbers (e.g. to identify the student), entering a sequence of digits to specify a sequence or set of response options rather than exactly one as an answer, and free text entry. When the latter becomes widely available, we can at last address a fundamental problem of discussion groups (such as research seminars) where many people want to ask a question: which is the best question to take for the group as a whole? Using only voice, we cannot know what the set of candidate questions is without having them asked. With textual group responses, everyone's questions could appear in front of the speaker and/or facilitator, and could then be grouped, sequenced, and sorted by priority. Meanwhile, as the technology (especially radio communication techniques) advance rapidly, we can focus on how we would use additional functions, and what their pedagogic rationale is.

Conclusion

The studies in prospect with this equipment should eventually allow us to pronounce on the validity of the design rationales presented in this paper. These studies will use measures of learning outcome, attitudes, and engagement as dependent (i.e. output) variables. They may range over, as independent (i.e. input) variables, two or more universities, three or more levels of university class and so student experience, two or more academic subjects, class sizes up to 300, and all the pedagogic strategy types described above.

References

Cohen, P.A. (1980). Effectiveness of student-rating feedback for improving college instruction: a meta-analysis. *Research in Higher Education, 13*, 321-341.

Draper, S.W. (1998). Niche-based success in CAL. *Computers and Education, 30*, 5-8

Draper, S.W., Brown, M.I., Henderson, F.P. & McAteer, E. (1996). Integrative evaluation: an emerging role for classroom studies of CAL. *Computers and Education, 26*(1-3), 17-32.

GRUMPS (2001). *The GRUMPS research project* [Online]. Available: http://grumps.dcs.gla.ac.uk/ [1 June 2001].

Hake, R.R. (1998a). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am.J.Physics, 66*(1), 64-74.

Hake, R.R. (1998b). Interactive-engagement methods in introductory mechanics courses. submitted to *J.of Physics Education Research*

Hoyet, H. (2000). Graphing Interactivity in technology-based training. *TechTrends, 44*(5), 26-31.

Landauer, T.K. (1995). *The trouble with computers: Usefulness, usability, and productivity.* Cambridge, MA: MIT press.

Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology.* London: Routledge.

Marton, F. (1981). Phenomenography - describing conceptions of the world around us. *Instructional Science. 10*, 177-200.

Marton, F., & Booth, S. (1997). *Learning and awareness.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mayo, E. (1933). *The human problems of an industrial civilization.* New York: MacMillan. ch.3.
Poulis, J., Massen, C., Robens, E., & Gilbert, M. (1998). Physics lecturing with audience paced feedback. *American Journal of Physics. 66*(5), 439-441.
Sims, R. (1997). Interactivity: A Forgotten Art?. *Computers in Human Behavior. 13*(2), 157-180.

Acknowledgements

This work is being supported in part by the EPSRC funded grant to GRUMPS (GR/N38114), and also in large part by the University of Glasgow, both directly and through the TLC project. Many thanks to Prof. Jim Boyle of Strathclyde University, whose rich existing experience with this equipment gives us confidence in proceeding, and many practical tips. We look forward to collaboration with him, and with David Nicol who has done a pioneering evaluation of Jim's work. Thanks also to Michael McCabe, Mae McSporran, and the anonymous referees for points that have been incorporated into this version.