

Increasing interactivity in lectures using an electronic voting system

S. W. Draper* & M. I. Brown†

Departments of *Psychology and †Computing Science, University of Glasgow, Glasgow, UK

Abstract

An overview of the experience of the opening two years of an institution-wide project in introducing electronic voting equipment for lectures is presented. Eight different departments and a wide range of group size (up to 300) saw some use. An important aspect of this is the organizational one of addressing the whole institution, rather than a narrower disciplinary base. The mobility of the equipment, the generality of the educational analysis, and the technical support provided contributed to this. Evaluations of each use identified (formatively) the weakest spots and the most common benefits, and also (summatively) showed that learners almost always saw this as providing a net benefit to them. Various empirical indications support the theoretical view that learning benefits depend upon putting the pedagogy (not the technology) at the focus of attention in each use. Perceived benefits tended to increase as lecturers became more experienced in exploiting the approach. The most promising pedagogical approaches appear to be Interactive Engagement (launching peer discussions), and Contingent Teaching – designing sessions not as fixed scripts but to zero in on using diagnostic questions on the points that the particular audience most needs on this occasion.

Keywords

campus, formative, interactivity, lecture, organizational change, portable, summative

Introduction

In this paper, we give an overview of our experience of introducing an electronic voting system for use in lectures. Our general hope was that the use of handsets would ‘engage’ the students: get their minds working and so prompt learning. The initiative includes an organizational change dimension: its eventual aim is not merely to look at the effect on learning in one context, but to explore both use and adoption across a whole university. In the initial two-year period, this was tried in eight different departments from psychology to computing science, from philosophy to

medicine, with first year and fourth year students, and in group sizes from 12 to 300. The amount of usage varied greatly from single-hour uses to routine use in a 20 lecture course, repeated and improved every year for three years. While some evaluation was carried out in nearly every case, the amount of evaluation was scaled to the usage. Furthermore, the kinds of benefit vary over a considerable range, depending not only on the subject matter being learned but still more also on the pedagogical aims and consequent methods of the lecturer. Thus this paper focusses on the overall aims and generalisable conclusions, rather than all the evidence in particular cases. More detailed papers on the more intensive uses either have been (Stuart & Brown 2003; Wit 2003) or will be published. Here, however, we focus on the overall lessons, which should be most informative both to those considering adapting this

Accepted: 26 January 2004

Correspondence: S. W. Draper, Department of Psychology, University of Glasgow, Glasgow G12 8QQ, UK.

Email: s.draper@psy.gla.ac.uk

approach to another context, and to the questions of what the general lessons might be, independent of particular subject domains or methods.

The project began from the theoretical motivation of addressing a general inadequacy of the lecture format (lack of interaction), and the technology was thus selected in the hope of being advantageous in all subjects. Its relationship to an older tradition of developing interactivity in lectures for introductory university mechanics, and the use of this equipment in that tradition, are discussed, as are the technological alternatives. Its use was evaluated systematically, including how it was valued by learners and by teachers, and its ease of adoption for the interested teachers. We then present the issues that are now our focus, especially that of attempting to move towards truly contingent teaching, where what is presented in a class depends more upon the audience's responses and not solely on a fixed script. Finally, we briefly discuss the relative importance of the technology and the pedagogical changes to the improvements in the learning experience observed.

It has already been shown (Crouch & Mazur 2001) that such a voting system when used as part of a particular pedagogical method (called 'peer instruction') in a particular disciplinary context (teaching first year mechanics classes in universities) produces large and statistically significant improvements in standardised test results. In their related change to a similar course Boyle and Nicol (2003), using the same equipment as reported on in this paper, also found a small upward shift in exam scores but with a pronounced effect in raising the 'tail' of the class (J. T. Boyle, personal communication). The questions that this project in effect addressed are: what are the prospects for generalising this? What is it like to try to roll this out from a very specific sub-disciplinary context to other subjects? Are there only gains from large changes to pedagogical approach, or are there some easy gains to be had from small changes?

Theoretical motivation for the intervention

In an earlier paper (Draper 1998), we argued that many applications of information and communication technology (ICT) to higher education had not been worthwhile, since while they generally managed to re-attain approximately similar standards of learning and

teaching, they did not demonstrably improve quality despite increased costs. The few exceptions that did improve quality were projects that had begun with identifying a significant deficiency in previous provision, and finding a way to use technology to address that real problem. We called this 'niche-based CAL' because these success stories did not offer general technological answers, but (as in evolution) were about the fit between each separate problem (or context, or niche) and the specific technical solution developed to address it. The conclusion was that to apply ICT to education successfully, pedagogy had to be put first, technology second.

An unstated implication of this was that we ourselves should look at the teaching practices around us, identify the weakest points, and try to discover how ICT could address these. One of the weakest points in the teaching at many universities is the use of lecturing, especially to large classes. The common diagnosis of what is weak in this method is the lack of interactivity. Teachers (i.e. lecturers) experience this as a feeling that they cannot get any discussion going and so lose much sense of how well the material is going over. This general diagnosis, and suggestions of remedies, is made in numerous places, for example Charman and Fullerton (1995), and Steinert and Snell (1999). An important case with a highly developed method of proved efficacy in teaching mechanics to undergraduates (Hake 1998) is that of 'Interactive Engagement', which we discuss later.

A more theoretical view is that because no overt response is required of students, little mental processing in fact takes place, and hence little learning, at least during the lecture. A technology aimed directly at this gap would be a voting system (for instance, one similar to that used in the 'Ask the audience' part of the TV show 'Who wants to be a millionaire?'), where every student can enter a response to a displayed multiple-choice question, (MCQ) and the aggregated results are immediately displayed to everyone. Because this argument applies to lectures in general, independent of subject, audience size or point in a degree programme, we obtained portable equipment that could be set up at any time in any lecture theatre, and advertised its availability in the university newsletter; the result was a wide variety of applications with teachers motivated to attempt this innovation.

In addition, an analysis was developed of ways in which this might be used pedagogically (Draper *et al.* 2002). Despite the limitation to an MCQ format, the following pedagogic uses seemed plausible:

- Assessment, both formative and summative e.g. practising exam questions with, in effect, instant self-marking and feedback.
- Formative feedback on learning (for teacher as well as learners). Self-assessment questions, which then show both each individual learner how well they know the material, and the teacher how well the class as a whole knows it.
- Formative feedback on teaching to the teacher. Anonymous feedback on questions about the delivery e.g. too fast, more examples wanted, etc.
- Peer assessment: where the class rates the performance of each learner's presentation or other work.
- Community mutual awareness building: whether a group of researchers at a workshop or a new first year class, a few minutes spent in asking and sharing some basic questions on where the participants are from (local, abroad), gender, age bands, other subjects they are taking, whether they view this subject with enthusiasm or trepidation can make everyone feel more at home, and more oriented within the group. Readers probably know this from experience at various meetings. It is equally thought to be important to university students in terms of academic and social integration (Tinto 1975).
- Experiments using human responses. In subjects such as psychology, effects can be not just described but demonstrated by collecting responses from the audience, and furthermore this allows each participant to experience their own personal response and to relate it to the group's mean and variability.
- To initiate a discussion, especially in small groups. Peer discussion, particularly of topics where peers disagree, is well known to be excellent in promoting conceptual advance. (Besides Piaget's and Vygotsky's foundational views, see Miyake (1986) and Howe (1991)). A good way to initiate this is to display a 'brain teaser' question, have the audience each select an answer, display the group disagreement (i.e. the broad spread of responses), and then without announcing the correct answer (if any), have the audience discuss it with their neighbours.

This then was the background for the introduction of the technology.

Alternative technical methods

The most basic requirement is for every student audience member to be able to vote, and a show of hands will do this. However, it is not very easy to see at a glance, in a group of say 200 people with randomly distributed and clumped votes, the difference between one third, one half, and two thirds of the audience. It is only possible to get one option at a time (not even the difference between yes, no, and abstain is possible with a single show of hands). Furthermore, this is not anonymous, and we have found that students quite often care about this. A system of 'flash cards' to be held up (Meltzer & Manivannan 1996), or a cube with different colours on each face that is rotated by the student to show their current vote, allows six rather than one option at a time to be registered, and feels somewhat more private to students (only near neighbours can easily see what an individual has voted), but the issue of adequate counting remains. Electric and electronic systems using installed wiring overcome these objections. These all have a set of buttons for each student, automatic and near-instantaneous counting, and the central display of the aggregated totals of votes. They have been sporadically used for some decades (Bridgman 1965; Appleby 1968; Cuban 2001). However, they seem to have been too expensive to maintain, and are fixed to a single room each. More recently, however, numerous infra-red systems and some radio ones have made mobile equipment possible, and reduced the overall cost.

The equipment

The particular equipment we adopted is known as personal response system (PRS), and was originally developed by Professor Nelson Cue, now of the Hong Kong University of Science and Technology. It consists of infrared handset transmitters (like half-size domestic TV remote controls) for each member of the audience, receivers linked in a chain to the serial port of a PC (usually a laptop) running the software, and displaying via a data projector to everyone in the room. Users know their 'votes' have been received when they see their handsets' unique ID numbers (given on a label on the back of each handset) appear on the main display.

Questions may be displayed on a separate overhead projector, on the same or a separate data projector (e.g. in Powerpoint), or simply announced orally. Questions may be prepared in advance, or on the spot, or both. In one case (giving Biology students a prelab tutorial), photomicrographs were displayed on one screen, text questions about them on a second, and the response registration feedback on a third.

With an audience of 200, it typically takes about two minutes to collect the answers; with 50 people, 30 seconds is enough. (Displaying and explaining the question is usually given additional time before that; for questions that require considerable thought by the audience extra thinking time during collection will be required, while discussing the answers is often given considerable time after collection.) When collection is stopped, a bar chart is projected showing the number or percentage of people who voted for each alternative. Thus everyone can see the degree of consensus, while each participant also knows what they themselves selected, and so how their response compares with the rest. Each person's response is anonymous to the rest of the audience, and also anonymous to the teacher, unless records are kept and used of which learner has which handset.

Enough equipment was acquired for use in our two largest lecture theatres simultaneously (350 and 300 seats), and we aimed to provide assistants to transport, set up, and take down the equipment. This frees the lecturer to concentrate on their job of managing the occasion as a whole, and allows the equipment to be used without having to book special rooms. For practised assistants (not always the case in the trials reported here), set-up of the receivers and laptop takes about three minutes. If, as is being increasingly done, receivers and cables are permanently installed in the room, some of this is avoided, but of course usually the computer still needs to be connected and started up. Handing out transmitters (handsets) is like distributing handouts: with organization it can be done as people enter the room, and similarly collected in boxes as they leave. Only one handset was lost using this arrangement in the first two years. Alternatively, as was done for one course, handsets can be given out, subject to a fine for non-return, to each student for a whole semester. This is a less efficient use of equipment, and resulted in about 5% not being recovered after the course, but a still greater problem was that on

a given day, 25–35% forgot to bring them. Both the hardware and software of the handset equipment have proved largely reliable. Unpractised assistants were occasionally a problem, causing delays. However, some of the biggest disruptions have been caused by not being able to get the data projection equipment provided by the university to function.

It is important never to allow the equipment to become the point around which the occasion is organised. It is perfectly possible and desirable to mix handset use with asking questions in other modes: by shows of hands, and simply asking the audience to volunteer answers. As lecturers become more fluent with the equipment, they tend to do this more and more naturally. The other modes tend to feel more spontaneous and quicker, while returning to the handsets gets every audience member (again) to contribute a response.

The equipment is in use elsewhere in the UK, in other countries such as the USA and Hong Kong, and in schools as well as universities. Here we report on the experience in one institution, based on evaluations in a variety of teaching situations, and with particular focus on its general applicability.

Uses over the first two years

Besides using the equipment in our own teaching, a short article was placed in the University newsletter (Draper 2001) and a talk given to an internal conference on teaching methods. These proved enough to elicit volunteers who had thought of 'niches': places where they believed that the equipment would substantially improve their teaching. Thus the trials in this startup phase have been with teachers who saw a definite value for their teaching and some particular way of using the equipment.

Between October 2001 and March 2003, it was used in Medicine, Dentistry, Veterinary Science, Biology, Psychology, Computing Science, Statistics, and Philosophy. There thus appear to be no constraints on the subject it can be used in. (In other UK universities we know of, it has also been used in French language, Economics (Elliott 2001), Mechanical Engineering (Boyle & Nicol 2003), and Mathematics.)

It was used with undergraduate levels from the first year to the final year. It was used on groups with sizes from about 20 to about 300. While it takes a bit longer

to collect responses from larger groups, it is still perfectly practicable with them, and in larger groups the expected added pedagogic value of the added interactivity is greater because it is more difficult to achieve there by other means.

It has been used for a range of periods. It was used on a one-off occasion to give medical students practice on an MCQ format exam, thus demonstrating that it can be used successfully by learners and teachers with no previous experience of it. (The success of this occasion led to it being used again with that class at later dates.) At the other end of the range, it has also been used throughout a Computing Science course with two lectures a week for a semester, and we now have data for three consecutive years of use there (Table 1).

The types of pedagogic use included:

- Exam practice (where the exam would be in MCQ format) i.e. formative assessment.
- Self-assessment questions: in effect formative feedback that tells each learner where they need to do more work.
- The same (type of) questions also give feedback to the teacher, and in some cases this was used to control what the teacher did e.g. spend longer on explanations on the topics where many got the wrong answer.
- Experiments (one trial of this here; but it has been more extensively carried out elsewhere).

- Community awareness building: in small amounts, and at open days.
- To launch discussion.

The evaluations

In most of the cases that the equipment was used an evaluation was performed. The evaluation approach was broadly that of Integrative Evaluation (Draper *et al.* 1996) and was concerned both with apparent problems, participant acceptance, and the extent to which the use of the equipment was or was not integrated into the teaching. Thus while it produced some summative results to inform decisions about whether to continue with the initiative, it was concerned with producing formative information: but mainly not formative of the equipment (the evaluations were not designed to help redesign the technology), nor even only of the pedagogic tactics, but of the delivery as a whole. This follows the view that learning outcomes depend on the ensemble of resources and activities presented to the learners and the integration of the technology with the others. In particular, information was gathered to help us improve our support service for lecturers, and to help each lecturer to improve immediately their teaching with the equipment. The amount of data sought had to be proportionate to how long the equipment was used. Thus in the longer uses learners were asked about some of the possible specific

Table 1. Teaching uses of handset equipment at the University of Glasgow 2001–3; there were some additional uses at workshops, open days, and conferences

Department	Class	Approx. number in class (not all attended session)	Number of sessions
Computing Science*	Level 1 2001–02	450	20 × 2
	Level 1 2002–03	300	20 × 2
Computing Science	Level 4	70	1
Psychology*	Level 4 Education	40	3
Psychology*	Level 4 HCI	30	8
Psychology*	Level 1	500	3 × 2
Philosophy*	Level 2 Logic	100	9
Philosophy*	Level 1 Mind and Body	260	1
Medicine*	Level 3	250	3
IBLS (Biology)*	Level 2	300	1 × 2
IBLS (Biology)*	Level 2	150	1
Veterinary Medicine*	Level 4	100	1
Dental School*	GPs (short course)	18	1
Medicine	Level 4	250	1
Statistics*	Level 1/Level 2	200	9

*Some evaluation carried out

advantages and disadvantages, while in the short, one-occasion uses only a single question might be asked.

Instruments used in this period included:

- Observation of lectures (with and without handset use).
- Informal discussions with students who had used the handsets.
- Use of the handsets to evaluate the use of the handsets i.e. simple questionnaires administered on the spot with the equipment.
- Written open-ended comments from students.
- Paper questionnaires.
- Discussions with lecturers before and after their handset use.
- Written feedback from lecturers after their handset use.

Formative issues

A major aspect of developing formative information was to identify the most important advantages and disadvantages as perceived by the learners. This was done in a sequence of three phases. Firstly, in an initial open-ended phase, interviews, informal observations and discussions, and open-ended feedback forms threw up a set of candidate issues. At this stage there was no information about how widely agreed, or how important any item was. Secondly, a list of nine of each type of these was presented back to some student groups, who were asked to rank order them, or simply to select the most important one. This is directly useful formative information, since it says which the most important issue is to address; but it does not directly give information on how important an item is (the worst feature of a very good system might still not matter much in balance against the advantages). One version of these lists is given below, and some examples of the top ranking issues in particular cases are mentioned. In the third phase, direct Lickert-style questions were asked about a few of the higher ranked of these items selected for relevance in that particular case: this phase could be used more summatively to address the issue of how worthwhile the initiative was, and eventually whether its value changed year on year.

Some benefits of using handsets to vote in lectures

1. Using handsets is fun and breaks up the lecture.
2. Makes lectures more interactive/interesting and involves the whole class.

3. I like contributing opinion to the lecture and it lets me see what others think about it too.
4. The anonymity allows students to answer without embarrassing themselves.
5. Gives me an idea of how I am doing in relation to the rest of the class.
6. Checks whether you are understanding it as well as you think you are.
7. Allows problem areas to be identified.
8. Lecturers can change what they do depending on what students are finding difficult.
9. Gives a measure of how well the lecturer is putting the ideas across.

Some problems with using the handsets in lectures

1. Setting up and use of handsets takes up too much time in lectures.
2. Can distract from the learning point entirely.
3. Sometimes it is not clear what I am supposed to be voting for.
4. Main focus of lecture seems to be on handset use and not on course content.
5. The questions sometimes seem to be for the benefit of the lecturer and future students and not us.
6. Annoying students who persist in pressing their buttons and cause problems for people trying to make an initial vote.
7. Not completely anonymous in some situations.
8. Some students could vote randomly and mislead the lecturer.
9. Sometimes the lecturer seems to be asking questions just for the sake of it.

All these items, good and bad, were perceived by at least some students in at least one of the applications, but often did not apply generally to all uses of the equipment. Furthermore, the relative importance of each item changed a lot over different cases. Generally speaking, the benefits stayed fairly stable while the disadvantages changed, as one would hope, as we improved our practice in the light of this formative evaluation data.

In the cases where advantages and disadvantages were asked about in detail, the highest scoring advantage across applications was most often 'checks whether you are understanding it as well as you think you are'. Anonymity ranked high in some classes, but in some others was not seen as important. In the first year computing science course in the second year of use one of the leading benefits was 'Makes me think

about the course material' and one of the leading disadvantages had become 'often don't have enough time to think before having to vote'. By the third year, this disadvantage was the only one that half the class agreed with (as opposed to being neutral or disagreeing), the next highest scoring disadvantage was only perceived as such by a third of the class. In contrast, there were three advantages, all of which were agreed to by at least 75% of the class: checking understanding, 'allowing lecturers and students to identify problem areas', and 'makes me think more about the course material'.

Overall benefit

One attempt at a summative question was to ask for a rating of degree of agreement with 'How useful do you think the handsets are'. This was asked every year in the computing class, and the rising rating year on year suggests, as the changing degree of agreement with advantages and disadvantages did, that the handsets were being used increasingly effectively in that class: see Fig. 1.

Apart from this role in year-on-year comparisons, while most agreed that the equipment was useful, this only shows that there are some perceived benefits, not whether on balance there is a net benefit and so that the

application is worthwhile. In later studies this question about net benefit was asked directly whenever possible: 'What was, for you, the balance of benefit vs. disadvantage from the use of the handsets in your lectures?' with the response options from 'definitely benefited' through neutral to 'definite negative net value'. This provides direct summative data, at least in terms of student perceptions, of the value of each application of the equipment. In all cases except one, there were a clear majority of students who reported that the advantages outweighed the disadvantages. For example, Fig. 2, which applies to the prolonged twice weekly use in computing, and Fig. 3 show the common pattern, while Fig. 4 includes the exception.

The exception, embarrassingly but perhaps significantly, was in a class given by one of the authors (Fig. 3: the HCI class). In this class, as with another class run by someone heavily involved in the whole initiative, there was a perception reported by a significant subset of students that the equipment was being used for its own sake, because of the enthusiasms of the teacher, rather than being of direct benefit to the class. This strongly reinforces the 'niche' argument, that only when education is put first do we actually see real benefits of technology. A second factor was that this class witnessed several problems

Fig. 1 Answers from three successive years (two parallel groups in the first year) in computing science to 'How useful do you think the handsets are?'

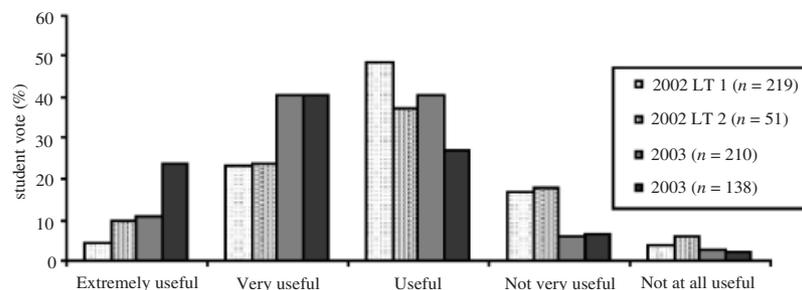
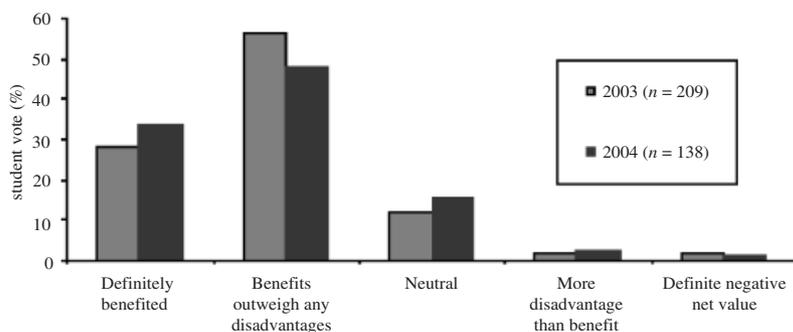


Fig. 2 Responses to the net value question in a first year computing science course in the second and third years of handset use: 'What was, for you, the balance of benefit vs. disadvantage from the use of the handsets in your lectures?'



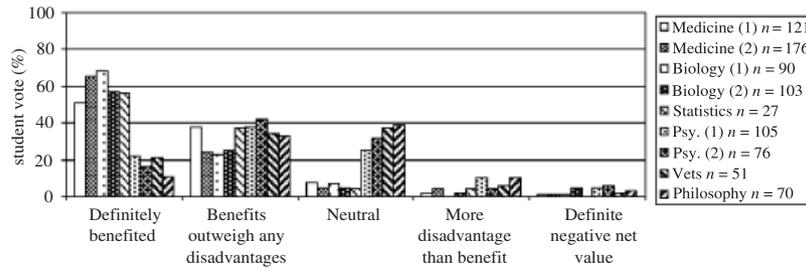


Fig. 3 Responses to the net value question in assorted other classes. The 'n' shown is the subset of the class present and responding at the time the evaluation question was put.

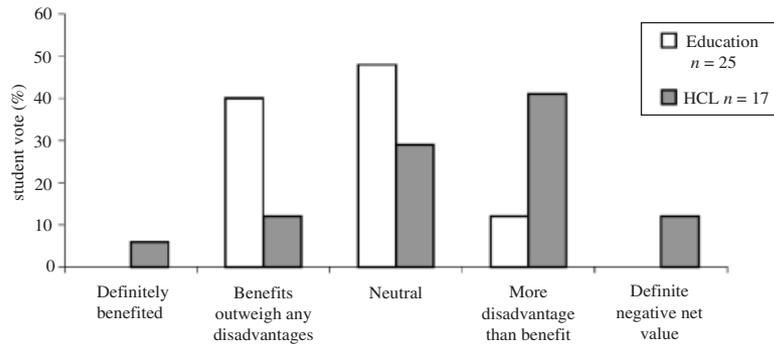


Fig. 4 Contrasting responses to the net value question in two different final year Psychology courses (first year of use).

with setting up the equipment, and so more of the disadvantages than in some other cases. A third interpretation is that this class (with typically about 20 attending on each occasion) already made use of many interactive techniques: participatory demonstrations of methods, buzz groups, structured discussions, and so on. Thus the relative increase in interactivity was less marked than in most classes.

Similarly all teachers except two felt that it had been worth it. A lecturer in the veterinary school gave this feedback: 'With the handsets, I could see exactly which points I had not conveyed clearly and could rectify it straight away, the major example being when I asked the students what I thought was a simple question – identifying the FCoV carrier cat! Although most (68%) got it right, but an astonishing number chose one of the other cats. I could see that they hadn't fully understood that many antibody positive cats are not infected. It was great, because the students who got the wrong answer are very likely the same ones who never utter a word in interactive lectures and it gave them a chance to participate anonymously.' The exceptions were one statistics lecturer from a group who had collectively decided to give the equipment a trial in a series of tutorials, and one medical lecturer from a group of three who had decided to use it in a joint session (with each taking a

turn to present). In both these cases, there was not only less personal commitment but also concern with time overall (would it be enough to 'get through' the planned material?) made worse by any use of extra equipment.

General conclusions from the first two years of use

In the light of these trial uses and the evaluation data from them, the evidence generally, so far at least, is that this equipment does provide a modest but worthwhile augmentation in the quality of the learning and teaching in lectures in the opinion of both learners and teachers. Furthermore, it does so in ways consistent with the theoretical reasoning that motivated its adoption, and even the exceptions so far can be understood in the same way.

It can be used successfully from the first session, at least by teachers with a prior idea of why and how it would help them in their particular situation, which amounts to knowing what questions they want to ask, and how that would fit into their lecture plan. Examples of this could be paraphrased as 'I want to give them practice at the MCQ exam format they will shortly be facing', 'I want to give them practice before their lab at interpreting photomicrographs', 'I want to

ask them which form of logical deduction they find most difficult, so I can concentrate the time on that', 'I want to drill them on classifying each of the evaluation instruments I'm teaching against a standard framework, so after describing each method I'll ask them how to classify it against each of the six dimensions in the framework'.

After the great majority of these uses (with the exceptions noted above), both learners and teachers judged its advantages outweighed its disadvantages.

In this startup phase, the equipment was used only by volunteers, and furthermore almost exclusively by those motivated by a clear pedagogic idea in relation to the equipment. What therefore is limited about this report of successful institutional change is that they were all keen, and all had their own spontaneous vision of how it would be a real help. What is good about this is that their visions were diverse, and they all worked first time in very varied contexts. Thus the expectation that it would be useful in all subjects is supported, but whether it could add value to all lectures is largely untested.

An issue that arises is whether this may be a novelty effect. Certainly, in almost all classes where it is introduced for the first time, there is a great rustle and smiles of interest. However, after a few minutes, and certainly by the end of the first lecture, the feedback showed that it was being judged again by whether it was serving a clear educational purpose, and if not, then complaints and down-ratings were articulated if asked for. Thus there is a novelty effect, but it seems to last only somewhere between 5 and 50 min, and this was true both when the audience was mainly female, final year Arts faculty students, and when it was mainly male, first year, computing students. In fact, our experience with this and other innovations is often the reverse of a novelty effect. On first introduction learners are ready to be sceptical (even if entertained), but in subsequent years take it for granted and rate it higher. This is probably due to two factors: the teachers becoming more practised and so using it better, and the learners being influenced by this growing confidence in the teachers. Thus innovation is often valued more not less highly as it becomes less novel, more familiar: an anti-novelty effect. Comparative ratings and teacher reports from the first year Computing Science class showed this pattern for the voting equipment across the two academic years.

The general motivation for trying this equipment was to introduce more interactivity into lectures. This issue was explored by asking students in some classes as to how likely they were to work out an answer if it was to be given in different ways: via the handsets, orally (with students putting up their hands), and so on. We first asked a handful of students informally after a variety of handset use lectures about this, and then carried out a systematic survey on this in some classes: see Table 2. Data from a philosophy logic class showed a similar pattern.

Continuing informal probes show that this remains a constant theme: asking questions via the handsets makes far more students actually think through and decide on an answer than presenting the question in other ways. Having to produce an answer oneself causes the mental processing; otherwise most students play the role of spectator and wait to see how it will be answered by others. This is strong support for the importance of one sense of interactivity: prompting mental processing in every learner's mind. This is true whether the question is one meant to provoke discussion, or simply a small problem of the kind that might form part of a basic test. Furthermore, students value this sense of interactivity, saying things such as 'how nice to be actually asked to think in a lecture'.

Another recurring theme is the importance of the anonymity provided by the equipment. In some ways this is surprising. It is easy to imagine that a first year student in a group of 300 people they don't know, faced by a lecturer they have no personal relationship with, is reluctant to answer in public. However, a group of about 30 (in final year psychology lectures on education) who knew each other well, and had shown no hesitation in joining in an oral group discussion previously, still said that the anonymity was important when challenged by a self-assessment question about the theory being taught that they felt uncertain about. Thus it seems it is not only about having a 'good' atmosphere within a group but also about how threatening each question separately feels. In contrast, it is very noticeable that few select the 'don't know' response option when that is offered in a handset question. Anonymity seems to function to induce people to pick a definite answer even when they are quite uncertain; and this in turn seems useful in getting people both to think in order to produce an answer, and then to take this (if they get it wrong) as a reason

Table 2. Likelihood of working out an answer for different response modes in a computing science course. This course was delivered by giving each lecture twice in a day (LT1 and LT2)

Question	% of students who voted for each option		
	2002 LT 1 219 students	2002 LT 2 51 students	2003 (LT1+LT2) 219 students
Given a problem to work out in a lecture, were you more likely to work out the answer if:			
Response option			
The class was asked for a verbal response to the question	3.7	0.0	0.5
The class was asked to vote on one or more answers using 'hands up'	1.8	7.8	0.5
The class was asked to vote on one or more answers using the handsets	32.4	15.7	40.6
None of the above (i.e. I never try to work out an answer)	4.1	5.9	2.7
All of the above (i.e. I always try to work out an answer)	20.6	21.6	17.8
'Verbal' and 'hands up' but not 'handsets'	1.8	2.0	2.3
'Verbal' and 'handsets' but not 'hands up'	7.3	3.9	3.2
'Hands up' and 'handsets' but not 'verbal'	25.1	35.3	28.3
No answer	3.2	7.9	4.1

for working on the point later. Thus anonymity seems often important (not just to break the ice and establish a good atmosphere at the start), and when mixed methods of interaction are used, returning to the handsets is probably still important because of this. This is one distinctive advantage of this equipment over other methods such as raising hands, holding up cardboard response cards, and so on.

The benefit most frequently mentioned by students was the feedback they got about their own understanding from many uses of the handsets. This supports the widely reported point that useful feedback is in short supply for students, and given the rise in student numbers, an evermore important bottleneck in UK university educational provision. Handsets are one way of providing immediate personal feedback to the whole class simultaneously (since they all know what they answered, and can compare that both to what is announced as the correct answer, and to what the rest of the class selected as the answer). In fact, this allows even relatively uninspired handset use to be valued by students. In one case, we persuaded a colleague to try using them, and although he agreed, he spent only a few minutes designing some simple self-assessment questions he tacked on to the end of his prepared lecture. Nevertheless, students regarded this as a worthwhile increase in value to them. In a different case, of statistics 'tutorials' of up to 200 students, students also sometimes showed no interaction in the overt, social sense: they resolutely declined all invitations to respond orally to questions, and in many cases didn't discuss questions with their neighbours

when invited to; but they still reported afterwards on how valuable they found the feedback provided on what they did and didn't understand correctly. This not only underlines the value of feedback to students and the potential of handsets to support this but also draws attention to the different senses of interactivity. Human-human interaction is one important kind of activity that facilitates learning, but it is by no means the only such kind supported by the handsets.

So in summary, in the applications in the first two years of introducing the equipment, the three most commonly important features of it that emerged, as reported by students, were: getting feedback to learners about whether they understand the material presented, that it does get most students to think about the question and decide on an answer while the alternatives do not, and that the anonymity is often important in achieving these benefits. What should be the next focus of interest in developing handset use?

Interactive engagement for teaching mechanics

An important body of work, predating the availability of the handset equipment, although now often using it, is that of 'interactive engagement' in teaching elementary mechanics (Newton's laws, etc.), mainly in first year university classes. The essential recipe is to introduce peer discussion within large classes, where the topic of discussion is a carefully designed MCQ of the 'brain teaser' type, i.e. a question that seems simple yet divides most audiences.

An example is as follows. 'Remember the old logo or advert for Levi's jeans that showed a pair of jeans being pulled apart by two teams of mules pulling in opposite directions. If one of the mule teams were sent away, and their leg of the jeans tied to a big tree instead, would the force (tension) in the jeans be: half, the same, or twice what it was with two mule teams?' Designing a really good brain teaser is not just about a good question, but about creating distractors, i.e. wrong but very tempting answers. In fact, they are really paradoxes, where there seem to be excellent reasons for each contradictory alternative. Such questions are ideal for starting discussions, although perhaps less than optimal as a fair diagnosis of knowledge.

Given a good question to provoke discussion, the exact classroom procedure to follow has been disputed (Dufresne *et al.* 1996; Mazur 1997). For instance, what is the best combination and sequence of individual, small group, and whole audience discussion with the teacher? This has been investigated empirically by Nicol and Boyle (2003).

In teaching mechanics, the interactive engagement approach has been shown to have a large positive effect, more than doubling the measured learning (Crouch & Mazur 2001), and to work across a large numbers of institutions, and so teachers (Hake 1998). It is natural to use handsets with this approach. The experience of Jim Boyle in Mechanical Engineering at Strathclyde University is that progressive reorganization of the teaching around this approach has effects on the timetable (prefer two-hour to one-hour sessions), the architecture (seating to facilitate small group discussion within large rooms with handset equipment built in), and the relationship of the teaching to the curriculum (abandon the commitment to cover all material in the sessions in favour of concentrating on the topics that are most difficult for that group). See for example Boyle and Nicol (2003).

This is certainly one important model to emulate, and a number of the trials discussed here did use questions to launch peer discussion. However, 'interactive engagement' as a general approach has been demonstrated only in one part of one subject area; and more importantly it seems to depend on a question bank of brain teasers, which each require a considerable effort to invent or discover. Furthermore, handset equipment supports many other kinds of pedagogic use than initiating peer discussion. While this techni-

que certainly warrants further development across subject areas, a slightly different issue has emerged as of general importance.

Contingent teaching

Students have widely and immediately shown that they value the feedback that handsets can be used to deliver on their own learning. The same questions and aggregated responses equally tell the teacher how well the class understands each issue. An increasingly interesting issue is whether lecturers can then respond to this information on the spot, by varying what they do. This may be called 'contingent teaching' by analogy to Wood *et al.*'s (1978) concept of contingent tutoring. The defining attribute is making teaching (the course of a lecture session) depend upon actions of the students, rather than being a fixed sequence predetermined by the teacher. To put it another way, this requires not just the students to interact (to be active and for that action to depend on the teacher and material), but the teacher to be interactive too. In the programming analogy, a lecturer must develop their lecture plans beyond the factory machine stage of executing a rigid, pre-planned sequence regardless of the circumstances, and discover the conditional statement so that they create a flexible branching plan, where which branches actually get presented depends on how the audience answers questions or otherwise shows their needs.

When using handsets, most teachers naturally do this in a small way by varying the amount of explanation of the question and alternative responses, cutting it short if most students gave the correct answer, expanding it if many got it wrong. However, most feel the pressure of a fixed agenda for the session to the point of preferring to 'finish' what they planned to 'cover' even in the face of evidence that they are failing to communicate its meaning. Clearly, the next stage is learning to design sessions that are more contingent. This is important because it makes the teaching relevant to actual needs. (It is in fact equally pointless to waste everyone's time in sessions that fail to achieve learner understanding, or in sessions on topics the whole group understands already.) This is more important than may be realised. Classes vary from year to year: lecturers with regular handset feedback report not being able to predict what learners

will find difficult from year to year. (Those who say they can typically do not in fact have much feedback about their students). It is also important because it cheers learners up enormously to see their response having a direct effect, and to see a teacher respond on the spot to their actual learning needs.

We are currently involved in three developments related to developing contingent teaching further. The first is to gain more experience in following the 'interactive engagement' direction of using the handsets as a tool to launch productive peer discussion. The second is a collection of technical developments in new software for the handsets by Quintin Cutts and others. Most of these new features are aimed at improving the extent to which teachers can examine and reflect on a classes' performance with a view to adapting the course from week to week and year to year, and linking what happened in one session to others (e.g. by displaying answer distributions from an earlier lecture beside those from the current lecture). The third is the adoption of the handsets by an innovative group of Statistics lecturers for use in giant 'tutorial' sessions with up to 200 students in a first year class (Wit 2003). These sessions are not for introducing new material, and so can focus entirely on how to meet the needs of the learners by attempting to adapt on the spot to what they needed. Rather than simply varying the amount of explanation for a fixed set of questions, they have begun to experiment with bringing a large diagnostic set of questions, and selecting the next question depending on the audience response to earlier questions. We have already written Web pages inspired by this (Draper 2003) offering advice to teachers developing materials for handsets. We hope to test these ideas more directly.

Interactivity

While interactive engagement and contingent teaching are two particular classroom procedures that can be employed with the voting equipment and have emerged as particularly promising, it is also worthwhile to consider a more general analysis in terms of increased interactivity, which probably represents the main reason for the broad success in many different cases that we have seen in the startup phase of our project.

These handsets are about learner and teacher interaction: about having what individuals think and do affect what others consequently think and do. The point of having individuals co-present is to allow this. In large groups, it is easy and efficient to have everyone hear what one person has to say (this aspect scales up very well), but a basic limitation is in awareness of what the many think or say. The voting technology directly addresses this fundamental problem in a limited yet largely effective way, which gets the issue of feedback from the many to scale up too. In educational applications, such electronic voting systems can be used to obtain immediate feedback to learners on their understanding and many first users concentrate on this, remedying one basic drawback to large class teaching. However, just as important, but requiring somewhat more care to achieve, is feedback to the teacher about how this particular group is coping. The dream of personal teaching is really about adaptive teaching: where what is done depends on the learner's current state of understanding. The handsets make this possible even with large groups, although it is additionally necessary for the teacher concerned to plan to do this, for instance by coming with a bank of diagnostic questions and other material and being prepared for alternative actions depending on the group response. Thus this equipment addresses an intrinsic weakness of situations with large groups of humans, and so has the potential for yielding real advantages over previous practices.

Note that a simple learner-centred or cognitive analysis only captures half of this, and is likely to miss the importance of teacher adaptation. The implied model of the pedagogical benefits is as follows, where '→' means 'enables' or 'facilitates' (i.e. is one prerequisite for).

Voting equipment use →
 MCQ design; pedagogical tactics →
 Interactivity →
 Learner's thinking and reflection

But also:

Voting equipment use →
 MCQ design; pedagogical tactics →
 Interactivity →
 Teacher's adaptation of what to present,
 in what detail →
 Improved learning, because adapted to
 learners' current need

Conclusion: the organizational dimension

The preceding sections show how this project raises interesting questions about how exactly learning benefits are obtained, and how these depend on particular pedagogic techniques that the equipment may facilitate. These questions would ideally be investigated by fixing on a single such technique and running experiments, preferably with random assignments to treatment groups, and with exam scores as an important outcome measure. While the work has illuminated these particular questions, it has also shown that a positive effect relatively independent of both subject matter and pedagogic method of use is generally obtainable, at least as measured by learner opinion on whether voting equipment was worthwhile in their case. This more general finding is important to the organizational dimension of the project that is a focus of this paper.

Our motivating theoretical analysis, our choice of mobile equipment and technical support, and our targeting of client lecturers across the institution were all part of the aim of showing that this technology can be beneficial across contexts and pedagogic methods, and of promoting the spread of this innovation. In terms of organizational change, early adopters were found and supported. Extensive Website materials for them were created (Draper 2003). We have recently conducted another round of promotion and attracted 24 more lecturers, registering serious interest and attending workshops. In terms of the sigmoid (S-shaped) curve of technology adoption, we may now be turning the lower corner into a phase of rapid rollout. It is thus time to reflect on where the expansion of this new practice and new technology currently stands.

Our sense of the first year of use is that in bracing ourselves against numerous little practical hitches, we were able to realise the pedagogical benefit we had anticipated, but were not immediately able to relax and reap still further benefits that were only just occurring to us. In the second year, with more confidence that it would all work without real anxiety (and function not just in a technical sense, but in the sense that sessions would go well and be well received and effective with students), we could begin to focus more attention on pedagogical benefits, and on the issue of how we might improve particular classes still further. Improved evaluations from the Computing Science class over the three academic years that we now have

results for there, together with the more confident feeling that lecturer reports, are one illustration of this. It seems reasonable to claim that immediate benefits for students have been achieved, and that in the first phase this was mainly due (as anticipated) to increased interactivity for the learners.

In summary, use of the handsets was judged by both learners and teachers to benefit them. It can immediately be used successfully by teachers new to it, provided (a) that they come with a niche-specific idea of how to use the equipment in their situation (although simply adding self-assessment questions seems to be valued almost generically), and (b) that there is human assistance sufficient such that no technical difficulties obtrude on the learning situation. Success is associated with increasing the interactivity of the occasion. Promising ways forward from the initial modes of use discussed here are to increase this interactivity by (a) peer discussion, and (b) more contingent teaching.

We began from the view that success depended on putting pedagogy first, technology second, and this was sharply reinforced at times in this project: firstly by the criticism surfacing in the evaluations when the students perceived the lecturers' motivations as being technology centered, and secondly by the further improvements evident as lecturer pedagogic technique improved with experience. Thus the benefit does not depend simply on the technology but on how well it is used on each occasion to promote, through learner interactivity or contingent teaching or both, thought and reflection in the learners. Thus even from an organizational change perspective, where the question can often seem to be 'what value does the technology have?', it is important not to lapse into thinking of it as a magic bullet but as more similar to central heating in contrast to individual fires. It is not essential to achieving learning or interactivity, which for millennia have been frequently achieved by other means; but nevertheless it makes the desired effect (the target room temperature in one case or interactivity in learning in the other) easier to achieve more often, in more contexts, and with much less effort and attention. Thus the requirement to keep teachers' minds on the pedagogy not the technological means may in the end be the main argument for the technology, while also showing how important seamlessness and effortless operation may be to its ultimate effectiveness.

Acknowledgements

The work reported here was partly funded by EPSRC grant GR/N38114/01 to the GRUMPS project, and partly by money for learning and teaching innovations from the University of Glasgow. We would also like to thank Professor Jim Boyle of Strathclyde University for sharing both his technical and pedagogical experience with us, and Professor Nelson Cue for his advice and help.

We are grateful to all the teachers who allowed us to observe their teaching even while it was developing, and both the students and teachers who gave us evaluation data. We would also like to thank Quintin Cutts and Chris Mitchell, our collaborators on this project throughout this period.

References

- Appleby E.C. (1968) Teaching aids and the practitioner. *Veterinary Record* **83**, 291–292.
- Boyle J.T. & Nicol D.J. (2003) Using classroom communication systems to support interaction and discussion in large class settings. *Association of Learning Technology Journal (ALT-J)* **11**, 43–57.
- Bridgman C.F. (1965) Innovations in the teaching of anatomy. *American Journal of Veterinary Research* **26**, 1552–1561.
- Charman D.J. & Fullerton H. (1995) Interactive lectures a case study in a geographical concepts course. *Journal of Geography in Higher Education* **19**, 41–55.
- Crouch C.H. & Mazur E. (2001) Peer instruction: ten years of experience and results. *American Journal of Physics* **69**, 970–977.
- Cuban L. (2001) *Oversold and Underused: Computers in the Classroom*. Harvard University Press, London.
- Draper S.W. (1998) Niche-based success in CAL. *Computers and Education* **30**, 5–8.
- Draper S.W. (2001) 'Want to try interactive handsets in your lectures?' *The Newsletter*, no.231 (The University of Glasgow) and as [WWW document] URL <http://www.psy.gla.ac.uk/~steve/ilig/ad.html> (visited 2003 May 1).
- Draper (2003) 'Interactive Lectures Interest Group' [WWW document] URL <http://www.psy.gla.ac.uk/~steve/ilig/> (visited 2004 Jan 21).
- 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**, 17–32.
- Draper S.W., Cargill J. & Cutts Q. (2002) Electronically enhanced classroom interaction. *Australian Journal of Educational Technology* **18**, 13–23.
- Dufresne R.J., Gerace W.J., Leonard W.J., Mestre J.P. & Wenk L. (1996) Classtalk: a classroom communication system for active learning. *Journal of Computing in Higher Education* **7**, 3–47.
- Elliott C. (2001) 'Case Study: Economics Lectures Using a Personal Response System' [WWW document] URL http://www.economics.ltsn.ac.uk/showcase/elliott_prs.htm (visited 2003 May 1).
- Hake R.R. (1998) Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* **66**, 64–74.
- Howe C.J. (1991) Explanatory concepts in physics: towards a principled evaluation of teaching materials. *Computers and Education* **17**, 73–80.
- Mazur E. (1997) *Peer Instruction: a User's Manual*. Prentice-Hall, NJ.
- Meltzer D.E. & Manivannan K. (1996) Promoting interactivity in physics lecture classes. *The Physics Teacher* **34**, 72–76.
- Miyake N. (1986) Constructive interaction and the iterative process of understanding. *Cognitive Science* **10**, 151–177.
- Nicol D.J. & Boyle J.T. (2003) Peer instruction versus class-wide discussion in large classes: a comparison of two interaction methods in the wired classroom. *Studies in Higher Education* **28**, 458–473.
- Steinert Y. & Snell L.S. (1999) Interactive lecturing: strategies for increasing participation in a large group presentations. *Medical Teacher* **21**, 37–42.
- Stuart S. & Brown M.I. (2003) Traditional and non-traditional resources: providing a well-supported learning environment. *Association of Learning Technology Journal (ALT-J)* **11**(3), 58–68.
- Tinto V. (1975) Dropout from Higher Education: a theoretical synthesis of recent research. *Review of Educational Research* **45**, 89–125.
- Wit E. (2003) Who wants to be... The use of a Personal Response System in statistics teaching. *MSOR Connections*, **3**, 5–11.
- Wood D., Wood H. & Middleton D. (1978) An experimental evaluation of four face-to-face teaching strategies. *International Journal of Behavioral Development* **1**, 131–147.