Summative project evaluation

A MANTCHI project report

(Deliverable, item number 15)

This report may also be found at http://www.psy.gla.ac.uk/~steve/mant/deliverable.html

by
Stephen W. Draper, Margaret I. Brown, & Sandra P. Foubister
Department of Psychology
University of Glasgow
Glasgow G12 8QQ U.K.
email: steve@psy.gla.ac.uk
WWW URL: http://www.psy.gla.ac.uk/~steve

For more information on the MANTCHI (Metropolitan Area Network Teaching of Computer-Human Interaction) project: see http://mantchi.use-of-mans.ac.uk/.

The MANTCHI project was carried out by Glasgow Caledonian University, Heriot-Watt University, Napier University, and the University of Glasgow, and funded by the Scottish Higher Education Funding Council (SHEFC) through the Use of MANs Initiative (UMI phase 2).

Preface

This report concerns the evaluation work done on the MANTCHI project. It does not set out to describe the project in general. Brief descriptions are found near the beginning of parts B and D. Further information should be sought in other papers and reports and through the project web site http://mantchi.use-of-mans.ac.uk/.

This report is mainly assembled from drafts of three other papers as they were on 24 August 1998, which comprise parts B, C, and D, and which may be published elsewhere. Those papers will be revised and improved in the near future, but there are no plans to update this report.

Contents

Part A: Introduction

This report is assembled mainly from other papers in order to provide a deliverable promised in the MANTCHI grant application, where it is described only as "Summative project evaluation" (item number 15). The meaning and scope of that now seems ambiguous between:

1) The evaluation work we did. Part B presents a large overview paper on the (integrative) evaluation work done on the project, which corresponds to reporting on what we did for project objective 2 (of 4): "To measure the educational effectiveness of this novel delivery of tutorial support using the method of Integrative Evaluation". One of the products of this evaluation work was a set of recommendations for delivering our teaching materials: these are presented in part C.

2) Evaluating our evaluation work. Part B also contains our self-criticisms of our evaluation method. We did what we promised, but what with hindsight might have been better evaluation aims?

3) Evaluating the project, not the educational work. This should surely be the content, not of one deliverable from part of the project team, but of the project final report. However as the project evolved and we perceived new needs, we did perform a study of a different kind which is important to an evaluation of the project itself: the cost benefit analysis of using our approach to collaborative tutorial teaching. This analysis is reproduced as part D. In addition, the conclusion (part E) offers a brief summary of the features of the project as a whole, its results, and their relationship of the evaluation studies.

The rest of this report therefore consists of:

- Part B: a long overview of the project's integrative evaluation work, based on 20 studies (listed in appendix 1). One output of this is part C.
- Part C: a list of the recommendations for delivering the ATOMs based on the evaluation studies.
- Part D: a cost benefit analysis of the MANTCHI approach to remote collaborative tutorial teaching based on an interview study of the teachers involved.
- Part E: A short overview of the relationship of these studies with the project as a whole, and its main aims, themes, and innovations.

Part B:

Evaluating remote collaborative tutorial teaching in MANTCHI

by
Stephen W. Draper & Margaret I. Brown

Introduction

This paper discusses the evaluation work on the MANTCHI project as a whole. The project focussed on "tutorial" material rather than primary exposition like lectures. It concerned teaching over a metropolitan area network, and so involved a mixture of face to face and distance learning, and separate issues of collaboration between learners in different places, and of teachers in different institutions. It was interested in the re-use of "tertiary" material such as past student solutions, and whether this was useful to learners.

The evaluation work comprised about 20 studies. Obviously full reporting of so many studies, even though of modest size, is not possible within a single paper. On the other hand, such a paper can (and this paper does) select the main conclusions to report, and draw on evidence across multiple studies to support them. It can also discuss evaluation methods and problems on a broader basis than a single study could.

The paper is divided into two parts, the first dealing with overall issues and the second with particular findings. The first part introduces the project's distinctive features and the particular demands these placed on evaluation. It then discusses our approach to evaluation, based on the method of Integrative Evaluation and so emphasising observations of learners engaged in learning on courses for university qualifications. The second part is organised around the major findings, briefly reporting the evidence for each. These findings are grouped into perspectives: learning effectiveness and quality, features of the teaching, and issues of the management of learning and teaching.

Part B1: The project and evaluation activities as a whole

The MANTCHI project

The MANTCHI project — Metropolitan Area Network Tutoring in Computer-Human Interaction — (MANTCHI; 1998), involving four universities in central Scotland for about 19 months, explored the development and delivery of tutorial material in the subject area of Human Computer Interaction (HCI) over the internet in existing university courses for credit. "Tutorial" was broadly defined to mean anything other than primary exposition (such as lectures). Typically this material is an exercise created at another site and backed up by a remote expert (often the author), who may give a video conference tutorial or give feedback on student work submitted and returned over the internet. In some cases students on different courses, as well as the teachers, interact. A unit of this material is called an "ATOM" (Autonomous Teaching Object in MANTCHI), and is typically designed as one week's work on a module for a student i.e. 8 to 10 hours, including contact time. Responsibility for the courses and assessment remained ultimately with the local deliverer. In this paper, we refer to these deliverers and also to the authors and remote experts as "teachers" (though their job titles might be lecturer, professor, teaching assistant etc.) in contrast to the "learners" i.e. university students.

Teaching and learning normally includes not only primary exposition (e.g. lectures) and re-expression by the learners (e.g. writing an essay), but some iterative interaction between teacher and learners (e.g. question and answer sessions in tutorials, or feedback on written work). Mayes (1995) classifies applications of learning technology into primary, secondary, and tertiary respectively by reference to those categories of activity. Technology such as email and video conferencing supports such tertiary applications. MANTCHI focussed on tertiary applications, and the additional research question of whether such interactions can usefully be captured, "canned", and later re-used. We called such canned material "TRAILs".

A key emerging feature of the project was its organisation around true reciprocal collaborative teaching. All of the four sites have authored material, and all four have received (delivered) material authored at other sites. Although originally planned simply as a fair way of dividing up the work, it has kept all project members crucially aware not just of the problems of authoring, but of what it is like to be delivering to one's own students (in real, for-credit courses) material that others have authored: a true users' perspective. This may be a unique feature. MANTCHI has in effect built users of higher education teaching material further into the design team by having each authoring site deliver material "not created here".

The project devoted substantial resources to evaluation of these teaching and learning activities. This paper offers a general report on that evaluation activity.

Our evaluation approach

The evaluation method we used was based on that of Integrative Evaluation (Draper et al. 1996), which was developed during the TILT project (Doughty et al., 1995; TILT, 1996). The main characteristic is to study teaching and learning in actual classroom use for reasons of validity that are of particular importance in the area of learning in Higher Education for two reasons. Firstly learning here is characterised by, and depends upon, conscious effort and choice, and hence on motivation: if you persuade subjects to use learning materials for an experiment they behave quite differently than if they are trying to get a qualification. Secondly, learning cannot be seen as a simple effect caused by teaching, but as the outcome of a whole ensemble of factors of which the intervention being studied (for instance some piece of learning technology, or in this case the ATOM materials) are only one. Trying to isolate one factor experimentally is usually unrealistic (leading to invalid studies) because in normal education, learners simply shift their use of resources in response to the particular characteristics of what they are offered. This view of teaching and learning is consistent with Laurillard's (1993) model of the process, which suggests that 12 activities (of which primary exposition such as a lecture or textbook is only one) are involved. When we measure learning outcomes, we are measuring the combined effect of all of these not just the one we varied, and the others are unlikely to have been held constant as learners adjust their use of activities and resources as they think necessary (e.g. ask questions of their tutor when but only when they feel the primary exposition was unclear). Such adjustment (i.e. this learner control) is a salient feature of higher education, and not something it is appropriate to suppress in a meaningful study, even if it were possible and ethical to do so.

This is bad news for simple summative evaluation that aims to compare alternative teaching to decide which is best, as Draper (1997b) argues. However evaluation is still possible and worthwhile, but turns out to be mainly useful to teachers in advising them on how to adjust the overall teaching to improve it: effectively this is formative evaluation of the overall teaching and delivery (not just of the learning technology or other intervention), called "integrative" because many of those adjustments are to do with making the different elements fit together better.

Consistent with that formative role, we have found that many of the most important findings in our studies have been surprises detected by open-ended measures, and not answers to questions we anticipated and so had designed comparable measures for. By "open-ended" we mean that the subject can respond by bringing up issues we did not explicitly ask about, so that we cannot tell how many in the sample care about that issue, since they are not all directly asked about it and required to respond. For example, we might ask "What was the worst problem in using this ATOM?" and one or two might say "getting the printer to work outside lab. hours". The opposite category of measure is "comparable", where all subjects are asked the same question and required to answer in terms of the same fixed response categories, which can then be directly compared across the whole sample. We use about half our evaluation effort on open-ended measures such as classroom observation and open-ended questions in interviews and in questionnaires, with the rest spent on comparable measures such as fixed-choice questions applied to the whole sample.

In previous applications of Integrative Evaluation, an important type of comparable measure has been either confidence logs or multiple choice quiz questions that were closely related to learning objectives in order to measure learning gains. In MANTCHI, while extensively used, they are of less central importance, as the material of interest is not the primary exposition but student exercises and the interactions and feedback associated with it. Furthermore, when we asked teachers why they used the exercises they did, their rationales seldom mentioned learning objectives but seemed to relate to aims for a "deeper" quality of learning. We tended to make greater use of resource questionnaires (Brown et al. ,1996) to ask students about the utility and usability of each available learning resource (including the new materials) both absolutely and relative to the other available resources. In this way, we adapted our method to this particular project to some extent, although as we discuss later, perhaps not to a sufficiently great extent.

Our evaluation work

We carried out about 20 studies in all (each of a different delivery of an exercise) in the four universities. These studies were divided into three phases. In the first period, we did some studies of the courses into which new material would be introduced, both to get some comparison for later reference, and to gain experience of studying tutorial exercises. In the second period (the autumn term of 1997), we studied the delivery of the first ATOMs, and circulated some initial practical lessons. In the third period, we studied many more deliveries of ATOMs.

A typical study would begin by eliciting information from the teacher (the local deliverer in charge of that course) about the nature of the course and students, and any particular aims and interests that teacher had that the evaluation might look at. It would include some classroom observation, and interviewing a sample of students at some point. The central measures were one or more questionnaires to the student sample, most importantly after the intervention at the point where it was thought the students could best judge (looking back) the utility of the resources. Thus for an exercise whose product was lecture notes, that point was the exam since lecture notes are probably most used for revision. For a more typical exercise where the students submitted solutions which were then marked, we sometimes chose the time when they submitted their work (best for asking what had been helpful in doing the exercise) and sometimes the time when they got back comments on their work (best for judging how helpful the feedback was). An example questionnaire from a single study is given in an appendix.

Part B2: Findings from the evaluation work

There are a number of different ways to organise our findings. The most obvious one would be to organise them by study i.e. evidence first then results. The disadvantage of that is that it tends to split up pieces of evidence that support each other if they were gathered in different studies. Instead, what seem to be the most important findings are simply stated, and then the evidence supporting those conclusions is summarised.

This gives a large set of small sections. One way of grouping them would be by stakeholder perspective: what did the learners think? what did the teachers think? what would educational managers e.g. heads of department think? The grouping adopted here is slightly different, into learning issues (e.g. what can we say about learning effectiveness and quality), teaching issues (e.g. is it worthwhile having remote experts?), and management issues (e.g. tips for organising the delivery of ATOMs).

There is some correlation between this grouping and the methodological one of comparable versus open-ended measures. Because we knew in advance we were interested in the issues of learning effectiveness we designed comparable measures for this, whereas most of the management issues emerged from open-ended measures (and complaints). However the correspondence is only approximate: for most issues there is some evidence from both kinds of measure.

Learning effectiveness and quality

The most important prior question about the project's teaching innovations is whether they increase or decrease learning quality and quantity. No significant or even noticeable differences were seen in exam and other assessment scores. However since these exercises were, like tutorials, only part of the teaching and learning activities on each topic, this neither is surprising nor would have been conclusive if observed.

Instead, we may ask questions about the value of the novel learning resources offered to students in this project, both generally and in comparison to others available for the same topic: were they valued or not, important or of little impact? These questions were mainly addressed through using forms of the resource questionnaire (Brown et al. 1996) which asks students to rate the utility of the learning resources available to them.

CSCLN

The clearest evidence for the value of an ATOM as a resource was found for the ATOM on CSCLN: computer supported cooperative lecture notes. In this ATOM, the class was divided evenly into teams, with one team assigned to each of the 20 lectures on that module plus one team for the index. Each team had to produce lecture notes for their assigned lecture in the form of a web page, structured as a set of key questions addressed by that lecture and answers for those questions, while the index page maintained a table to these other pages both by timetable (when the lecture was given) and by question (merged from the content of the pages). (This ATOM, and the pages produced by students, may be seen on the WWW at Draper; 1998.)

The main evidence came from a short questionnaire which, since lecture notes find their main use when revision for exams is being done, was administered directly after the exam. Of 59 students, 98% responded; and of these 84% said they had referred to the communal lecture notes, 76% said they found them useful, and most important of all, 69% said they found them worth the effort of creating their share of them. They also, as a group, rated these web notes as the third most useful resource (after past exam questions and solutions, and the course handouts). This shows that, while not the most important resource for students, nor universally approved by them, this exercise had a beneficial cost-benefit tradeoff in the view of more than two thirds of the learners.

The same ATOM delivered in different universities
Another good source of evidence comes from the UAN (User Action Notation) ATOM which was delivered to four groups of students at three universities. (This ATOM may be found through the project web pages; MANTCHI, 1998.) All students were asked "How much did you benefit by taking part [in the UAN ATOM exercise]?". Only 3 students in one university actually rated this as zero benefit. They were also asked the more directly interesting question "How did you rate the ATOM as a method of learning compared to the 'traditionally' delivered units of work experienced on your course?". At the first university, where the ATOM's author was also the local deliverer, 41% rated the ATOM as a superior, 50% as a similar, and only 9% as an inferior method. At the second, only 25% rated as superior and 75% as inferior. At the third, there was a very low response rate (8%) for the questionnaire with this question, but in that sample all (100%) rated it as a superior method.

This is clearly a mixed story. The unfavourable responses are clearly associated in the data as a whole with a high number of complaints about delivery (rather than content) issues, which are discussed below, and also with it not being directly assessed or compulsory. An interesting point here, though, is that the method was most favourably received (at least for the UAN material) where the author was also the local deliverer: so that everything was constant except the formatting as an ATOM. This is direct comparative evidence on the ATOM format itself.

Three ATOMs at one university
Three different ATOMs were delivered in a single course (along with non-ATOM-ised topics) at one university, which should have allowed comparisons to be drawn by the same students among ATOMs, and between ATOMs and other topics not organised in this way. However the main feature here turned out to be the declining numbers of students completing the ATOMs. In this case, the ATOMs were not done for direct credit (marks given for coursework), but only indirectly as the ATOMs were topics that would later be assessed for credit in other ways. Many of these students stated that it was not worth the effort of doing the work without direct credit. In a class of 50, 37 did the first ATOM, 15 the second, and none did the third. This of course destroyed the opportunity for a nice cross-ATOM comparison, but directed our attention to the workload question.

Students were asked "Was the 'workload' of the ATOM right for you?" on a 5 point scale. (This question appears in the appendix.) It is interesting that at a university where the ATOM was compulsory, only 32% of the students (who were in year 3 of 4 undergraduate years) rated it above (harder) than the neutral point, whereas at the university where it was not directly assessed (and the learners were a mixture of year 4 and M.Sc. students) 62% rated it as harder work than seemed right.

Features of the teaching

Remote experts
A feature of many of the ATOMs was the involvement of a remote expert at another university. For one ATOM, students at three universities were asked about the usefulness of "receiving feedback etc. from the remote expert on your group's solutions to the tasks". On a 5 point scale with the lowest point meaning of no use at all, the proportion rating it at one of the top 3 ("useful" or better) points were 87%, 57% and 75% in the three universities. This combines the usefulness of getting the feedback with the fact that this came from a remote rather than local expert. It supports the idea that remoteness is at the least not an important drawback for this function.

Open-ended measures suggest a mixed story. For instance at one university a class contained both year 4 undergraduates and M.Sc. students. The latter perceived much more benefit in having a remote expert than the former. Elsewhere, some students suggested that a benefit of remote experts was not that they had more authority (as national or international experts in the topic), but that because they were not in charge of assigning marks, the students felt freer to argue with them and challenge their judgements. This, if generally felt, would certainly be an advantage in many teachers' eyes, as promoting student discussion is often felt to be difficult. It is also interesting in that it is largely the opposite of what the teachers' seemed to feel. For them, the remote expert gave them confidence to deliver material they did not feel a deep grasp of, and to handle novel objections and proposed solutions that students come up with.

Tertiary materials

In some ATOMs, "tertiary" materials (i.e. past exercises, student solutions, and tutor feedback on those solutions) were made available to students. Exploring the use of this kind of material was one of the original aims of the project. Obviously it could not be provided until the ATOM to which it belonged had already been delivered, and so had generated student material for later re-use (unless simulated by the ATOM's author). Evidence of the utility of such material (called "TRAILs") to later students is positive but scanty.

When the statechart ATOM was delivered at one university, there were 50 in the class, of which 24 completed the questionnaire while 15 did the exercise, but only 9 did both. Of these, 6 used the TRAIL, and all 6 of these found it at least "useful". When the same ATOM was delivered at a second university, there were 11 in the class, of which 9 returned the questionnaire while 6 did the exercise as well as the questionnaire. Of these, 2 used the TRAIL and both of these found it useful. Thus although we may say that 100% of those who used a TRAIL rated it a useful resource, the numbers using it (whether from choice or simply from happening to notice it) were too low to give much certainty about this positive result by themselves. Open-ended comments were a second source of evidence supporting the positive interpretation, although on an even slenderer numerical base: "[They] Gave an indication of what was expected, though we felt the quality of the submissions was generally poor, we had no knowledge of the acceptable standard required.", and "Bad examples more useful than good. Can see how (and why) NOT to do things. This is much better than being told how to do something 'this way' just 'because'."

Still other kinds of evidence also suggest that this is an important resource to develop further. Firstly, theoretical considerations on the importance of feedback for learning support it. Secondly, in the CSCLN ATOM, the most valued resource overall was past exam questions and outline answers: a similar resource to TRAILs. Thirdly, and perhaps most important, in courses without such resources the open ended responses frequently ask for more feedback on work, model answers, and so on, strongly suggesting a widespread felt need for resources of this kind.

Management issues

From the start of our evaluations, as has been the case in many other projects, many of the points that emerged as problems were not about learning outcomes nor about the design of the learning material itself, but were practical points about the management or administration of the activities (e.g. informing students properly about resources and deadlines, availability of computing and other resources). In some descriptions of the educational process these issues are called delivery or implementation (cf. Reigeluth; 1983). From our perspective of seeing learning as the outcome of a whole set of activities (not the one-way delivery of material), we categorise these issues as the management of the learning and teaching process: about coordinating and organising those activities, rather than designing their content. This view is presented as an extension to the Laurillard model in Draper (1997a), and seen as at bottom a process of negotiation (tacit or explicit) between teachers and learners.

These findings did not mainly emerge from comparable measures designed to test learning outcomes, but usually from open-ended measures that yield (among other things) complaints by students: mainly open-ended questions in questionnaires adminstered to whole classes, interviews with a subset of nearly every class we studied, and the direct classroom observations we did in a majority of our studies. In the next part of this report (part C) we present these findings, together with suggestions for responses. Full lists of them, usually with the student comments transcribed in full, were fed back to the course deliverers for use in improving delivery next time.

Conclusions on the evaluation performed within MANTCHI

What we didn't do

The most obvious question to ask about the project is whether our approach (say, the use of ATOMs) was a good thing? Did it work? Do ATOMs cause learning? The direct way to answer that would be to run experimental groups that got them, and other groups that did not, or got a different presentation of the same topics, and measure learning outcomes in the same way in all groups. We did not do this because it is unethical not to teach students we have been paid to teach, because in most cases students would have found a way to learn assessed topics in any case from other resources especially as our ATOMs were not themselves the primary expository resource, because if (to avoid those problems) we studied students who did not have to learn the topics for credit we would be studying a situation far removed from the one of interest (motivation is the single key factor controlling learning in higher education, since students are responsible for their own learning), and because in any case no modern theory of education regards material as simply causing learning. We therefore remain content not to have attempted a direct, controlled, comparative study.

Nevertheless we were alert for any negative summative evidence, such as a sudden slump in exam results, which if found, would certainly signal a problem requiring attention. No significant change in exam results emerged. Our own measures of learning outcomes were mainly in students' self-estimates (confidence) about topics. But although these measures indicated satisfactory learning outcomes, they could not be compared across years to get ATOM vs. non-ATOM measures, as a main consequence of the project was to introduce new curriculum content.

What we did wrong

Given that our focus was tutorial exercises, we should perhaps have developed an outcome measurement method especially for these. In evaluating primary (expository) material, the method is usually to elicit learning aims and objectives, and to test the extent to which these are achieved by each student i.e. to use measures directly related to these (Draper et al.; 1996) such as multiple choice question quizzes. Early in the project we collected something analogous to learning objectives by asking each teacher why they used each exercise: we called these Tutorial Provision Rationales (TPRs). It was immediately striking that teachers did not usually give testable learning objectives as the purpose for these exercises, but broader aims. For instance "... the most important part of learning occurs by actually doing the big exercise... I feel the 'intrinsic feedback', i.e. the direct experiences of what does and doesn't seem to work, is the main source of feedback." "by applying the technique they are going to be thinking about HCI evaluation (initially) and design (later) from a new perspective. Also working with a common language for HCI should help to provide insights." "...They are encouraged to try and verbalise why they like a system .... This enables them to realise they need a vocabulary to discuss these issues."

We failed to devise measures, as perhaps we should, to test whether the aims expressed in the TPRs had been achieved. This might have been a more appropriate summative measure in this or any project about tutorial material than the usual measures of learning objective achievement such as quiz and exam results, which focus on facts and specific skills, not depth of understanding or strength of belief or appreciation of a problem.

Even given our evaluation designs, we had some significant problems in executing them. With hindsight, this is most obvious in our missing the opportunities for good comparisons for one ATOM delivered in several universities and for several ATOMs delivered in a single course at a single university. We missed making the most of these opportunities by failing to ensure that most or all the students took the ATOMs (increasing numbers opted out in some courses), and even where students did the ATOM work, we sometimes failed to get them to complete the evaluation instruments.

Allowing students the option of whether to do the ATOM, either explicitly or implicitly by awarding no direct assessment credit for the exercises, seems liberal but undermines the potential for learning from the innovation. It is probably a mistake to view this as giving students a free choice and perhaps inferring from the low uptake that the material was inferior to other material. In fact, the assignment of credit by teachers seems to be one of the most powerful signals (as perceived by students) about what is important: no credit implies that the teacher doesn't value it. On the other hand, the impulse not to force students to take new material reflects the wisdom of experience in introducing innovations. Arbitrary funding constraints set this project at less than two years. A more sensible duration would be three years, with the first spent in preparations and studying the "status ante", the second in introducing the innovations in a rather uncontrolled way in order to catch and remedy the main problems while minimising the risk to those students' interests, and the third in a more controlled and uniform delivery. Our actions in our second and final project year turned out to be a compromise between the plans, both desirable, appropriate to a second and third year.

The failure to get a good response rate to data collection in many cases is again an issue related to the unwillingness of the teachers to impose pressure on their students. In this project, as in others, we found teachers to be staunch defenders of their students against the rapacious desire of the evaluators to extract huge amounts of data regardless of the danger of alienating students by exhausting them. Furthermore, possibly less creditably, teachers are reluctant to sacrifice "their" contact time to having students (for example) fill in a questionnaire, even though this is essentially an activity in which the teacher learns from the students about the state of their shared teaching and learning activity. Yet the use of such scheduled time for evaluation did turn out to be important. Our attempts to use lab. times were not very effective as only a minority of students on these courses would be present at any one time. Our attempts to use a WWW questionnaire met with a very low response rate, apparently mainly because the students did not have "processing web pages" as a regular activity, especially not one that involved the definite work of filling in a questionnaire. Email questionnaires, though somewhat better since students often did have replying to their email as a regular activity, still brought a fairly low response rate. As sales people know, personal contact and an immediate deadline (however artificial) both greatly enhance the response rate. Hence, while the reasonable interests of students must be considered, time at scheduled class meetings does seem a requirement of effective data gathering. An interesting case was the gathering of crucial data for the CSCLN ATOM at an exam. Even the evaluators were worried that this would cause protest, but in fact almost complete compliance was obtained without any objection, though the voluntary nature of the exercise was stressed. But on close consideration, it can be seen that the request was in proportion (5 minutes of questionnaire filling tacked on to 2 hours of exam), it was at a time when stress was relieved (after, not before, the exam) and when the subjects had no other urgent engagement to go to, and it had the personal element (the invigilator was the teacher who wanted the data and was making the appeal face to face).

With hindsight, then, we can see what went wrong and what we could have done better. Different project members exhibited different strengths, and it is probably no coincidence that the one who got the best evaluation data did the least in delivering ATOMs authored by others and setting up potentially interesting comparisons, while those who did the most at the latter were least effective in organising the extraction of a good data set. The project was effective in collaborating in producing material and in exchanging it for delivery in other institutions, but that collaboration should have been pushed through to a greater extent in planning the evaluations. Without effective data gathering, much of the other work loses its value. In effect, the evaluators had a plan and designed satisfactory instruments in every case, but sometimes failed to secure enough active cooperation to reap the benefits.

Integrative evaluation is, in contrast to controlled experiments, organised to a great extent around the teachers delivering the material. This is because it aims to study delivery in real teaching situations. The shortcomings in data gathering in this project show that despite that orientation, it is necessary to secure some concessions from teachers for the benefit of the evaluation goals in the form of scheduled class time for evaluation and of ensuring that students do use the material by giving credit for it in order to learn from the innovations. While evaluation can and should be planned with the teachers and around their constraints, we do have to insist to a greater extent than we did in this project on formulating a definite evaluation plan in each case capable of leading to definite findings, pointing out that otherwise much of the effort of the project is wasted. In fact what is not hard — and should be done — is to enlist student sympathy by presenting the aims of the whole intervention, how their data will affect the quality of future teaching, and the whole plan complete with the evaluation actions they will be asked to cooperate with.

What we did right

Nevertheless, despite those reservations, this was probably not an important failing in this project. Our evaluation seems satisfactory in the formative purpose central to integrative evaluation of discovering problems for the delivering teachers to remedy. A sample of these were given in in an earlier section. They are typically to do with how to adjust the delivery to improve its effect. While some were detected early enough to influence later delivery, the project was too short for many of them to have been acted on yet.

Failing to get better comparative data on learning outcomes is probably of little importance. All the indications we have suggest that the quality of learning from ATOMs is at least as high as from other similar material (i.e. learning quality per topic was at least maintained on average), but that there was no increase in quality so large as to make that the important feature of the project. Instead, the real gains lie elsewhere: in the introduction of topics that the local deliverers of a course judge desirable but would otherwise not teach. In other words, the gains are in curriculum content, not in how a fixed curriculum is presented. Evaluation based on student responses is important to check that quality is at least maintained, but is unlikely to yield evidence about the value of changing the curriculum. The evaluation we did, while it could be improved, seems adequate for the summative purpose of quality checking, as well as for the formative purpose of guiding improvements in the overall delivery of the teaching.

What we need to do in future

As the project unfolded, and its main benefits gradually emerged, it eventually became clear that future improvements to its evaluation should fall into two classes: firstly, the improvements to method discussed above (direct measures of the attainment of the aims expressed in TPRs, and more complete planning of data gathering), and secondly completely new forms of evaluation, not foreseen when the project began, to address out new perception that the main benefits are in facilitating curriculum change by supporting teachers in adopting on to their courses material (ATOMs) authored elsewhere on topics they are not already familiar with. We would therefore propose at least three further types of evaluation:

1. Evaluating the benefit of curriculum change. Presumably this would require surveys of teachers, professional bodies, and employers on the relative utility of various topics. A further approach to this could be to study large student projects (on courses which have them), and analyse what topics and techniques are actually used by and useful to students when they are engaged in HCI projects.

2. Formative evaluation, not of student use of the materials, but of teacher use of the materials. That is, the project concerns use of material by those other than its authors. This adoption is a crucial task performed by teachers, the difficulty of which will depend to a significant extent on the form in which the material is offered, and the support if any offered by the original authors. Formative evaluation of this adoption task can improve formats and support tools. We are just beginning work of this kind as we evolve the format of ATOMs with a view to dissemination beyond the project.

3. Whether teachers decide to adopt an ATOM, at least outside the project and any special motivation that supplies, will depend upon their perception of the costs and benefits such adoption will afford them. This is crucial to the longer term success of this work, but requires a quite different kind of evaluation: centered on teachers not learners, and requiring an investigation into what they count as benefits, and the development of methods for measuring costs as experienced by teachers. As we began to recognise the need for this late in the project, we launched a small interview study of the teachers involved, asking how much work (time) it had taken to author ATOMs, to deliver them, whether they thought they would continue to use the ATOMs beyond the project, and why. This will be reported elsewhere. One crucial view that emerged was the idea that the main benefit for some of using someone else's ATOM was to be able to cover a topic the deliverer thought important but was not confident in teaching (a curriculum gain), but that after having the author as a remote expert for a year or two, the deliverer would probably be confident in delivering it solo in future. This would mean that ATOMs were a form of staff development more than a permanent form of collaborative teaching.

Conclusions

Overall, the evaluation studies provided formative information for improving the overall delivery of the material, and summative evidence on learning outcomes and quality that suggests that the ATOM-ised materials are of at least as high a quality as other material, although by no means always preferred. One telling piece of favourable evidence was that on one course students preferred the ATOM authored by their teacher to other parts of the course also developed by him.

Our studies suggest however that the main gains are in improved curriculum content and in staff development (expanding the range of topics a teacher is confident of delivering), but that different evaluation methods must be developed and applied to study that properly.

So with hindsight we should have spent less effort on learner evaluation and more on teacher evaluation: on whether they felt their work was better or worse, and more or less difficult. Cost measures would be a crucial part of this, as they will determine whether the project is carried forward: i.e. would the teachers use ATOM material without the extra motivation of participating in a funded project?

Part C:
Lessons on delivering tutorial teaching and ATOMs

by
Margaret I. Brown & Stephen W. Draper

This report presents collected lessons, findings and recommendations, learned during the MANTCHI (1998) project as a result of 20 or so evaluation studies, concerning tutorials and how to deliver web-based tutorial support (ATOMs) to students on HCI courses in 4 Universities. (An overview of their findings is given in another report: Draper & Brown, 1998.)

This report is structured into the following sections:

C1&2. The lessons (tutorials and delivery of ATOMs): if you want to know what we recommend, just read these.
C3. The basis: a short discussion of the kinds of evidence underlying this report, which you should read if you wonder just how much faith to put in them.
C4. The theoretical view: a short discussion of the kind of lessons these are.

C1. The lessons: Tutorials

This section presents our collection of findings and recommendations from the evaluation of HCI tutorials before the introduction of the MANTCHI ATOM structure for tutorial material.

1. Expectations
Access to past exam papers, specimen questions and worked examples gives students an idea of the approach they should take to learning and processing the lecture and other material. This is very highly valued by many students.

2. Practical experience
Students reported the importance of practical experience of actual interfaces, exercises, examples etc. and considered that they required more of these on their courses along with more practical experience of "new technology". Some students studying several formalisms suggested applying the different formalisms to the same interactive device.

3. Feedback
Students valued feedback and considered small tutorial groups were ideal for this. Even without the expected feedback, many still valued the practical experience of exercises.

4. Collaboration
Students were on the whole enthusiastic about collaborating with students in other universities. Those who had been involved in the first MANTCHI collaborations identified some of the benefits (seeing /hearing other students' experiences) and disadvantages (having to be available at the same time as the other students) of collaborating.

5. Information about the students involved
Students at different levels/courses may have different "requirements" and may require different kinds of tutorial support.

6. Video conferences
Students who had been involved in a video conference, considered that these should only be held for a specific, well defined purpose. Technical problems can interfere with a conference especially if lecturers had not experienced video conferencing before.

C2. Lessons: Delivery of ATOMs

This section presents our collection of findings and recommendations from the delivery of ATOMs

1. Computer and other Technical Support
The Network is not 100% reliable. Adequate technology has to be available and working to deliver/support an ATOM. Lecturers delivering ATOMs have to have alternative plans in hand in case the Network goes down. This is one reason for providing paper-based resources. If the ATOM requires students to use other equipment (in the case of the two ATOMs: scanners and pdf readers), these have to be accessible as do sufficient numbers of suitable computers.

2. Web-based vs. Paper Resources
Web-based instructions and resources may also need to be given to students on paper. During student use of some ATOMs, lecturers handed out paper-based instructions and resources. In some cases this was because the students were unable to access the web-based resources, in other cases it was because the lecturer wished to give the students additional instructions which superseded those on the ATOM web page.

Students usually download and print the web-based resources which is less efficient than these resources being centrally copied on to paper and handed out. Students reported that though it can be useful to access information etc. electronically, this is not always possible and anyway they like having a hard copy that they can make notes on. This also covers the problem of the network not functioning when needed by the students. It is also likely that the students will not have continuous access to computers while completing their assignments.

3. Information about the students involved
Knowledge of the students' previous experience is useful to lecturers involved in collaborative teaching before lecturing/conducting a remote tutorial. Local teachers need to brief remote teachers on this.

4. Assessment of ATOM tasks
If the ATOM tasks are not directly assessed, students may not complete the tasks. Where courses contain more than one ATOM and the tasks are not directly assessed, students may be less likely to complete the tasks for the later ATOMs. If the ATOM tasks are not directly assessed, students are more likely to report that the workload is too heavy than if it is directly assessed.

5. Students' Expectations
Expectations should be clear. ATOMs may involve remote experts, web-based instructions, and learning resources as well as some "in house" lectures, handouts and other resources. Students require information about what is available to them and what is expected of them in the way of self-tuition (resource-based learning) etc.

6. Content of ATOM
It should be clear if local instructions about the assignments (completion, submission etc.) differ from those on the ATOM home page. The ATOM (or the course web page) should contain clear information on: which resources will be delivered locally (in house), what to use: (e.g. a real physical radio alarm in an exercise on formal descriptions); access passwords; the date solutions should be submitted; exactly how solutions should be submitted; the approximate date on which web-based feedback will become available.

7. Time
Instructions about the ATOM resources and assignments have to be sent to students in plenty of time. Students do not all check their e-mail every day. Students admitted that even if they are given information in plenty of time they may not act on it. However where web-based resources (or any resources) have to be used before an assignment is to be attempted, students have to be given clear instructions in plenty of time for them to be able to plan and use the resources. They have to have the information to allow them to manage their time effectively.

8. Remote Expert and Local Deliverer
It should be clear to students whether the "in-house" teachers are "experts" or "facilitators". Each ATOM has a domain expert. The lecturer delivering the ATOM to his/her students need not be an expert in the subject. It is useful if the students are made aware that the lecturer may be "facilitating" rather than "teaching" and also that the work will involve "resource-based" learning utilising the ATOM web-based resources and a domain expert.

9. Feedback from Domain Expert
Students should be alerted when the web-based feedback on their solutions is available. They should also be alerted when feedback on the solutions from other universities is available. I.e. posting them on the web without e-mailing an announcement is unsatisfactory.

10. ATOMs involving Group Work
Group work involves extra organisation and time which has to be taken into account. Students recognised the benefits of group work, but found that it took more time than working in pairs or alone. This appeared to matter more where the task was not directly assessed. If possible, group work should be mainly within regular timetabled sessions of course to avoid clashes between courses. Similarly video conferences should also be within regular timetabled sessions. (The general problem is that of organising group meetings and irregular class meetings, which suddenly require new times to be found in the face of, for many students, conflicting classes and paid employment.)

11. New Types of Resources
Students may need to be encouraged to access and use new types of resources
Students varied in their use and reaction to the resources available on an ATOM. Many students did not use the TRAILs and other solutions and feedback. Until they become familiar with such resources they may need to be encouraged to use them. However we do have some evidence of the resources including solutions and feedback (tertiary resources) being re-used by some students while completing their projects/ essays.

12. HyperNews Discussion Forum
In future it may be necessary to manage the discussion in some way as the discussion forum was hardly used. During the use of the first two ATOMs this was really just used as a notice board for submitting solutions and getting feedback from the "Remote Expert". One student who did not attend the ATOM lab/tutorial sessions reported using the solutions and feedback on Statecharts and ERMIA to learn/understand these formalisms. In later ATOMs, solutions were submitted on web pages.

13. Collaboration between Students from different Universities
Rivalry between students at different Universities can result from ATOM use. Although this can be a good thing, we have to be careful to avoid the collaborations from discouraging some students from actively participating. Collaboration is mainly perceived as a benefit by students but on one of the ATOMs involving students at two Universities, comments from students at both universities indicated some rivalry and annoyance at comparisons used in the feedback.

14. The Integration of ATOMs into Courses
ATOMs are discrete units. The point has been raised that ATOMs could fragment a course, reducing the possibility of relating that topic to other parts of the course. This could be a problem especially if several are used. It is something that should be kept in mind. Integration of the ATOMs may be improved by asking students to write a report involving the topics studied on the ATOMs used, as this appeared to be successful with some referring back to the solutions and feedback.

C3. The basis: what is the nature of the evidence

The evaluations on which these recommendations are based were carried out in three phases. Those in Phase 1 led to the findings reported in Section A and those in Phases 2 and 3 led to the lessons in Section B. We won't give the actual evidence for every lesson for reasons of space and your boredom. Here is a discussion of the kind of evidence, and as an illustration all the evidence underlying one of the findings. This should be enough for you to understand the degree of strength and weakness of the evidence, and so to estimate the degree of belief merited by these lessons.

These findings did not mainly emerge from comparable measures designed to test learning outcomes, but usually from open-ended measures that yield (among other things) complaints by students: mainly open-ended questions in questionnaires administered to whole classes, interviews with a subset of nearly every class we studied, and the direct classroom observations we did in a majority of our studies. Full lists of the lessons, usually with the student comments transcribed in full, were fed back to the course deliverers for use in improving delivery next time. For one example item, we give details of the evidence on which the finding and recommendations were based. Should it be important to clarify an issue first identified by open-ended measures, then a more systematic measure can be applied. For instance, when the difficulties of group-work, and claims about high work load appeared, we then designed some systematic measures of these to investigate them further. Similarly, should one of the lessons in this report be particularly important to you, then you should include some specific measures of it in your own evaluations.

An example of the evidence
Here is some of the evidence on which the lesson "Web-based vs. Paper Resources" (no.2 in section B) was based.

In one study, all were asked if they had any problems while accessing the web-based resources. 25% reported some problem, examples being "Password problems plus early setbacks with software.", "On learning space — crashes". In a second study, all students were asked "Did you experience any difficulty gaining access to any resources / activities during the use of the ... ATOM?" 3 (13.6%) reported problems: "Remote web page" "server was down from where I had to access on-line." "Lab was too busy during lab sessions". They were also asked about resources for which there was insufficient time, which yielded comments including these: "Remote web page was too remote, took a very long time to view", but another student said "None! Most are web-based and therefore can be accessed at any time, when most convenient". In a third study, students were asked "What else would have helped at the two tutorials this week?" which elicited an 83% response rate including this long reply: "Computer equipment that worked! A lot of time was wasted in tutorials trying to fight with the equipment being used. It is not a necessity to teach through the use of computers when teaching to a computer course. In fact the opposite is true because computer students above all recognise the problems that can occur by over complicating a problem by using advanced computing e.g. the newsgroup on a web site (where a simple newsgroup added on to [the] news server would have achieved the same inter-communication and been far more reliable/faster than web browsing) and using the scanner (where simply drawing the chart on the computer would have been much faster and produced much clearer results for everyone to view). This is not a criticism of the ATOMs or the teaching method but more of the implementation which although seeming perfectly reasonable proved only to hinder our progress in learning about this topic!" In a fourth study, students were asked "Did you print the ATOM information and scenario from the Web?"; 10 (45.5%) said yes, 12 (54.5%) said no. They were asked if they used the paper or web form: 7 (31.8%) said paper, 10 (45.5%) said web, 4 (18.2%) said both. They were asked to explain why; among the numerous comments were "I like to save paper", "I took the work home", "some documents don't print well", "Web-based was easier to refer to related documents because of links". In a fifth study, printed versions were provided but students were asked if they had already printed out the web documents: 25% said yes. When asked which form they used, 45.8% used paper form, 20.8% the web form, 12.5% used both, and 20.8% didn't answer. In a sixth study, when asked how the ATOM compared to traditionally delivered units, one student said "Personally, I do not like using the net as a learning aid, I spend enough time working on a PC as it is without having to rely on the World Wide Wait to scroll through text on screen. Call me old fashioned, but I do prefer reading from books/journals/papers - a bit more portable and quicker to access - I wish I'd recorded how much time I waste during a week logging on, waiting for Win95 to start, waiting for Netscape etc. etc. etc. If I have an hour free in between lectures it is just impractical to get any work done on a PC." In a seventh study, when asked to comment on "How useful do you consider the ... ATOM Web-based resources were to you in learning & understanding ...? ", two explanatory comments (for low usefulness ratings) were "Items in pdf format prohibited many people viewing the docs", and "Paper-based notes are easier to manage and access. Paper notes don't crash!"

C4. The theoretical view

The main product of integrative evaluation (the method we used: Draper et al., 1996) is formative recommendations, not mainly of the software or other material being tested, but of the overall teaching delivery of which it is (only) one part. From the start of our evaluations in MANTCHI, as has been the case in many other projects, many of the points that emerged as problems were not about learning outcomes nor about the design of the learning material itself, but were practical points about the management or administration of the activities (e.g. informing students properly about resources and deadlines, availability of computing and other resources). It is this set of problems, and recommendations for avoiding them, that are presented in this report.

In some descriptions of the educational process these issues are called delivery or implementation (cf. Reigeluth; 1983). From our perspective of seeing learning as the outcome of a whole set of activities (not the one-way delivery of material), we categorise these issues as the management of the learning and teaching process: about co-ordinating and organising those activities, rather than designing their content. This view is presented as an extension to the Laurillard model in Draper (1997), and seen as at bottom a process of negotiation (tacit or explicit) between teachers and learners.

Many of the lessons in this report may seem obvious to readers, not so much from hindsight but because they are familiar points in the education literature. They are often rather less familiar to higher education teachers (who seldom read that literature), and who have very many such practical details to deal with in delivering any course (another reason for calling them "management" issues). This suggests that many gains in learning and teaching quality might be made, not by technical and pedagogical innovation, but by attention to best practice at this management level, backed by integrative evaluation to detect and feed back those points that emerge strongly as issues in each particular case.

Part D:
A cost-benefit analysis of remote collaborative tutorial teaching

by
Stephen W. Draper & Sandra P. Foubister

Introduction

A strong tendency in the evaluation of learning innovations and learning technology is towards learner-centred evaluation. There are at least two important reasons for this: if the aim is to establish claims about educational effectiveness, this is most directly done by measuring learning outcomes (i.e. studying the performance of learners) rather than only the opinions of teachers or other experts; and if the aim is to discover the often unexpected bottlenecks determining performance, which often vary widely between different cases, then again observing real learners in their actual situation is crucial. Our own work on Integrative Evaluation (Draper et al., 1996) has been in this direction. However there are other important issues which cannot be addressed in this way, but instead require some form of teacher-centred study. That is particularly the case with innovations whose main benefit is likely to be in saving costs in some form.

As with studying learning benefits, the issues are likely to be complex and the literature is much less developed (but see, for example, Doughty, 1979; Doughty, 1996a, 1996b, and Reeves, 1990). Identifying the kinds of benefit (and disbenefit), some of them unanticipated, is at least as important as taking measurements of those kinds that were expected. This paper reports an attempt at this, based on interviewing 10 teachers involved in an innovative project on remote collaborative tutorial teaching.

The MANTCHI project

The MANTCHI project — Metropolitan Area Network Tutoring in Computer-Human Interaction — (MANTCHI; 1998), explored remote collaborative task-based tutorial teaching over the internet. The project involved four universities in central Scotland for about a year and a half, and worked in the subject area of Human Computer Interaction (HCI). The material was delivered in existing university courses for credit. "Tutorial" was broadly defined to mean anything other than primary exposition (such as lectures). Typically this material is an exercise created at another site and available as web pages. Sometimes it is designed to use a remote expert (often the author) as part of each delivery, who may give a video conference tutorial or give feedback on student work submitted and returned over the internet. In some cases students on different courses, as well as the teachers, interact. A unit of this material is called an "ATOM" (Autonomous Teaching Object in MANTCHI), and is typically designed as one week's work on a module for a student i.e. 8 to 10 hours, including contact time. Responsibility for the courses and assessment remained ultimately with the local deliverer. In this paper, we refer to these deliverers and also to the authors and remote experts by their role as "teachers" (though their job titles might be lecturer, professor, teaching assistant etc.) in contrast to the learners, who in this study were all university "students".

A key emerging feature of the project was its organisation around true reciprocal collaborative teaching. All of the four sites have authored material, and all four have delivered material authored at other sites. Although originally planned simply as a fair way of dividing up the work, it has kept all project members crucially aware not just of the problems of authoring, but of what it is like to be delivering to one's own students (in real, for-credit courses) material that others have authored: a true users' perspective. This may be a unique feature. MANTCHI has in effect built users of teaching material further into the design team by having each authoring site deliver material "not created here". It is also a system of collaborative teaching based on barter. This goes a long way to avoiding organisational issues of paying for services. However the details may not always be straightforward, and the future will depend upon whether costs and benefits balance favourably for everyone: the core reason for the present study.

Evaluation of learning in MANTCHI

There was extensive evaluation work within MANTCHI, reported elsewhere, of the educational effectiveness of this material in actual classroom use, based on the method of Integrative Evaluation (Draper et al. 1996). Overall, the evaluation studies provided formative information for improving the overall delivery of the material, and some summative evidence on learning outcomes and quality that suggests that the ATOM-ised materials are of at least as high a quality as other material, although by no means always preferred by students. One telling piece of favourable evidence was that on one course many students preferred the ATOM authored by their teacher to other parts of the course also developed by him. Our studies suggested however that the main gains are in improved curriculum content and in staff development (expanding the range of topics a teacher is confident of delivering), something that our learner-centred evaluation could not directly demonstrate, but that teacher-centred evaluation could.

Another thing those evaluation studies could not show was whether we could expect teachers to use the ATOM materials beyond the end of the project and its special motivations for the participants. The educational benefits seem to be sufficient to warrant this, but not enough to provide an overwhelming reason by themselves regardless of costs and other issues. This would depend upon whether teachers found them of overall benefit. Another kind of study was needed to investigate these issues.

The purpose of the study

Whether teachers decide to adopt an ATOM, at least outside the project and any special motivation that supplies, will depend upon their perception of the costs and benefits such adoption will afford. This is crucial to the longer term success of our work, but requires a quite different kind of evaluation: centred on teachers not learners, and requiring an investigation into what they count as benefits, and the development of methods for measuring costs as experienced by teachers. As we began to recognise the need for this, we launched a small interview study of the teachers involved, asking how much work (time) it had taken to author ATOMs, to deliver them, whether they thought they would continue to use the ATOMs beyond the project, and why. This yielded information on estimated quantity of work, on the kinds of cost and benefit perceived, and also on where the ultimate significance of this ATOM approach (and hence of the MANTCHI project) might lie.

Method

We accordingly performed a short study in the final month of the project, consisting of retrospective interviews with the participating teachers (authors, remote experts, local deliverers). Each interview lasted about an hour. The agenda for the interviews was to ask how much time and effort had gone into activities related to creating and delivering the ATOMs, whether that teacher expected to use the ATOMs beyond the end of the project, and to identify what they thought the costs and benefits were. Actual times were obtained (and are shown in the table below), although the accuracy of these estimates is probably quite poor due to the retrospective nature of the measure: this discussion took place well after the actual delivery of the ATOMs, in some cases over a year later. At least as important a result was the insight afforded into how these teachers think about the pros and cons, the costs and benefits, of these materials i.e. identifying the kinds, rather than the quantity, of costs and benefits found to be relevant by the participants. We also noted comments about how to use ATOMs, the relative advantages of each type, the best way to use a remote expert, etc., although not all are discussed here.

The variety of ATOMs

ATOMs vary considerably in their nature, which of course affects the costs associated with each. This section briefly describes those features likely to affect time costs for teachers, and hence the values given in the table. (The ATOMs themselves will be available via the project website: MANTCHI, 1998.)

From this point of view of time costs, there are three types of ATOM. In the first group described below, a remote expert was actively involved in the delivery and this was a cost not incurred in other ATOMs. In the second group, students interacted between institutions, incurring extra coordination costs, and implying that deliverers at both institutions were involved simultaneously. In the third group, there was no dependency during a delivery on people at a remote site (although there were dependencies on remote resources such as web sites). This grouping is only for the purpose of bringing out types of cost; in a classification in terms of pedagogical method, for instance, the CSCLN and remote presentation ATOMs would be grouped together as being based on teachback.

The CBL (computer based learning) evaluation ATOM concerned teaching students how to perform an educational evaluation of a piece of CBL. The students had to design and execute such an evaluation, and write a report that was assessed by the local deliverers. The interaction with the remote expert was by two video conferences, plus some discussion over the internet (email and a web based discussion tools). The UAN, ERMIA, and Statechart ATOMs each concern a different notation for specifying the design of a user interface. These ATOMs each revolved around an exercise where students, working in groups, had to specify an interface in that notation. These solutions were submitted over the internet, marked by the remote expert, and returned over the internet.

The CSCW ATOM is a group exercise, which involves students working in teams assembled across two institutions to evaluate an asynchronous collaboration tool (BSCW is suggested). They first work together using the tool to produce a short report related to their courses, and then produce evaluation reports on the effectiveness of the tool. The formative evaluation ATOM takes advantage of the fact that students at one university are required to produce a multi-media project as part of their course. In this ATOM, students from a second university are each assigned to one such project student, and perform an evaluation of that project, interacting with its author remotely through email, NetMeeting, and if possible video conferencing. The ATOM on remote student presentations is a form of seminar based course, where students take turns to make a presentation to other students on their reading, and are assessed on this. In this ATOM, these presentations are both to the rest of their class and, via video conference, to another class at another university.

The CSCLN (Computer Supported Cooperative Lecture Notes) ATOM required students to create lecture notes on the web that accumulate into a shared resource for the whole class, with one team assigned to each lecture. There was no role for a remote expert. The website evaluation ATOM involves the study and evaluation of three web sites on the basis of HCI and content. Students complete the exercise on their own, over the course of a week, and submit and discuss their evaluations via Hypernews. In the website design ATOM students work in groups to produce a web site; there is no remote collaboration. In this ATOM, as in fact in all the ATOMs in this group, the exercise could be reorganised to involve groups split across sites, as in the previous group.

Result table

Here is a table that allows comparison between the time estimates of the various elements in an ATOM's lifecycle. Since many of the ATOMs were elaborations on pre-existing tutorial exercises, the presence or absence of an existing version is also noted.

ATOM	Pre-existed	Finding resources	Authoring	Preparation - *Author*	Preparation - *Deliverer*	Delivery	Marking	Revision
CBL	YES	-	3 - 4 hrs	2 hrs	3 hrs	2 X 2hrs video-conference 1 hr photo-copying	1 hr per group	0
UAN	YES	-	4 hrs	0.5 hr	**Napier:** 2 days **Heriot-Watt:** 1 hr	0	0.5 hr per group + 0.5 hr feedback on Web	0.5 hr
ERMIA	YES	24 hrs		4 - 7 hrs	0	Normal tutorial time	1 hr per group	4 - 7 hrs
Statecharts	NO	16 hrs	4 hrs	-	**First delivery:** 2 days **Subsequently:** 0	3 hrs	20 mins/ group 0.5 hr printing + 1 hr put feedback on Web	4 hrs + 8 RA hrs
CSCW	YES	-	2 hrs	1 - 1.5 hrs	0	1.5 hrs	Normal marking time	0
Formative Eval	YES	-	1 hr	13 - 14 hrs		2.5 hrs	3.5 hr	0
Remote Present	YES	-	(No ATOM description)	**Napier:** 4.5 hrs + programmer time **Glasgow Caledonian:** 1 hr + programmer time		4 X 1hr	0	0
CSCLN	NO	-	3 - 4 hrs	-	4 hrs	0	0	4 hrs
Website Eval	YES	-	0.5 hr	0.5 hr	0	(ATOM not used)		0.5 hr
Website Design	NO	16 hrs	2 hrs	0	0	(ATOM not used)		8 RA hrs

N.B. in the case of the ERMIA ATOM, the interviewee gave a single time (24 hours) for finding resources and authoring combined.

Accuracy of times

The times are probably underestimates (one subject said at the end of the interview "I bet these are all underestimates"). One of the authors, having been interviewed as a subject, later found a rough time diary he had forgotten about. Comparing the times recorded in that diary with the estimates he gave in the interview 4 months later shows he had lowered his estimate by about 30% for both the ATOMs he was involved in. The diary itself may err on the low side by missing a few occasions, and by being at best filled in retrospectively at the end of the day.

A second problem is that of accuracy in the sense of comparability (not systematic underestimation) of the times given. Many respondents noted that it is very hard to estimate "time" in this context. The time it takes to do something, such as physically type in an ATOM description, may not have much relationship with elapsed time — from, say, the original outline of the ATOM to the final, usable, version. People also mentioned "thinking" time and "research" time, e.g. "Do I count an hour in the bath spent thinking about it? An hour at home? An hour in the office?", and "I regard that as a free by-product of research thinking!". Nevertheless, every person interviewed felt able to give rough estimates. These vary very widely, e.g. from 0.5 hour to 24 hours for writing an ATOM description, so it may be that the implicit definition of "time" was indeed different between respondents.

The research aspect of MANTCHI will have increased the time costs because of the need to monitor various details, both at the time and in retrospect. This "costing" exercise has added yet a further hour for each of the dozen or so lecturers involved.

Cost categories used in the table

The columns in the table represent the main categories of time cost that emerged from the interviews.

- The "Pre-existed" column records whether there was an existing exercise on which to base the ATOM.
- "Finding resources" refers to collecting material for the author e.g. papers for the primary exposition.
- "Authoring" is writing the ATOM pages themselves.
- "Preparation — Author" is time the author or remote expert spent preparing for an interaction (e.g. a video conference).
- "Preparation — Deliverer" is similar work by the local deliverer.
- "Delivery" is contact time with students by local and remote teachers, and any other delivery costs (e.g. collecting student work).
- "Marking" is generating feedback on student work.
- "Revision" is rewriting the ATOM material in the light of earlier deliveries for re-use in a new delivery.

Authoring from scratch often requires little creativity or design time, as people only volunteer for ATOMs they already know how to write, or perhaps already have written in some form. The "authoring" category therefore is mainly about writing, and not about designing or creative thinking. In general it might be useful to have a "design" or "creativity" category, but in this study it would usually have had small quantities in it.

Authoring in general often has a significant element of iterative design to it, meaning that first draft authoring is often quite cheap, but we need to allow for the cost of revisions after a delivery or two, as these are really part of the process. Thus the authoring column in the table represents an estimate for a teacher considering joining in an ATOM exchange, but should not be compared directly with authoring times in other media (such as textbook writing) where some revisions would be part of the author's work. Conversely, the "revision" column combines (confounds) revision work that increases the quality with revision work simply to adapt the written material for a new occasion (e.g. replacing times and places in handouts, modifying URL links). N.B. the times for the two ATOMs not yet delivered are estimates of the latter time for adaptation. Future work should distinguish these two types of revision.

In summary, in general authoring might have four related categories: creativity or design, collecting resources to be used, actual writing, revision of the content in the light of the material having been used.

Time for marking is proportional to the number of solutions marked. The table gives the time per solution, which must be multiplied by the number of solutions in any attempt to predict times for future deliveries. This is one of the biggest issues in agreeing an exchange involving a remote expert: if class sizes are very different, marking loads will be different. There are several possible ways forward: different group sizes might mean the same number of solutions to mark even with different class sizes; a remote expert might just mark a sample with exemplary feedback, leaving the rest of the marking and commenting to be done by the students and/or local deliverer.

A related point to note is the use of groupwork in most of the ATOMs, since each group only produces one solution to be marked: thus the fewer the groups, the less time needed for marking. However larger groups certainly make it harder for students to agree meeting times, and may be less effective in promoting individual learning. Thus there is probably a strong tradeoff between learning quality and teacher time costs here, which is even more obvious in cases where the student work never gets marked at all or the feedback is of low quality.

Kinds of cost

One of the central results of this study is the identification of categories of cost, which future studies might design special instruments for. Each of the column headings above are one such category. As noted, "revision" should be dived into two, and perhaps "authoring" should be split into creative design and writing, besides one kind of revision and collecting resources.

One issue that emerged from the interviews was that time is not a currency with a fixed value. Time spent or saved at a period when other pressures on time are high is more valuable than at times of low pressure. Thus being able to serve as a remote expert at a low pressure time in return for getting the services of a remote expert at a high pressure time could be a very "profitable" trade ("The ATOM fell at a time when I was very busy, so having X do the marking was very useful"); while the converse could be disadvantageous even if the durations involved seem to balance. ATOMs move time costs around the calendar, which may be good or bad for the participants.

Another issue is also important: how difficult an activity feels to the teacher (perhaps measurable in terms of their confidence about performing the activity). The fundamental advantage to the trade behind ATOMs is that a remote expert usually feels it is easy to field student questions on the topic (or comment on unusual solutions), while the local deliverer would feel anxious about it. The time spent in each case might be the same, but the subjective effort, as well as the probable educational quality, would be significantly different.

A kind of cost not visible in this study is the groundwork of understanding what an ATOM is. For project members, this was done at a series of project meetings and perhaps in thinking about the ATOMs they authored, and is probably missing from all the time estimates. If new teachers were to use the ATOM materials, this learning of the background ideas might be a cost. On the other hand, this is comparable to the costs of any purchaser in gaining the information from which to decide whether to buy: a real cost, but often not considered. With ATOMs, a teacher would have to learn about ATOMs before making the decision whether to "buy in" at all, rather than while they were using them.

Finally, there were clearly costs in using some of the technology e.g. setting up video conferences, getting CSCW tools to work. As noted below, much of this can be written off as a cost of giving students personal experience of the technology as is appropriate in the subject area of HCI. This would not apply if the ATOM approach were transferred to another subject area, but retaining those communication methods. However it is also true that such costs will probably rapidly reduce as both equipment and staff familiarity with it improves.

Kinds of benefit

A number of kinds of benefit, although with no quantitative estimates, are apparent.

1. Subjective effort (or confidence): donating time at teaching you find easy in exchange for support at teaching you would find difficult.
2. Prime time vs. time at low pressure periods: this exchange could be a cost or benefit.
3. Using an exercise you did not have to write, in exchange for an exercise you wrote for your own use anyway. You may have to write it more carefully for use at multiple sites, so the gain is then: not having to write a new exercise in exchange for improving one you have written already.
4. The main value may be in a better curriculum. It is clear that even within the project, local deliverers only "took" exercises they felt would improve their courses (there were more offers of ATOMs at the planning meetings than takers i.e. the topics developed were more demand-driven than supply-driven). We can say definitely that all deliverers believed they improved their courses (although sometimes this may have been offset to some extent by the costs of moving topics around) by improving the selection of topics or the depth in which they were treated, although we do not have another measure of curricular quality.
5. Some teachers expressed the idea that after having delivered an ATOM in someone else's speciality a few times, they would probably feel comfortable delivering it without a remote expert. This would mean

that ATOMs were serving the function of staff development, rather than remaining permanently as an item of exchange.

6. In some ATOMs, the students were given a personal experience (rather than only an abstract concept) of the topic e.g. experience of CSCW by collaborating with a remote group via software. That is, the ATOMs additionally amounted to practical laboratory exercises i.e. a learning objective in themselves. This additional benefit comes from the synergy between the HCI content area being taught and the project's exploratory use of learning technology. Thus for instance in the CSCLN ATOM, collaborative lecture notes would probably benefit any course, but the students (and teachers) were also particularly keen on this as an occasion to practise web authoring for its own sake.

7. In the ATOMs involving students at remote sites, this contact was often seen as a benefit in itself: seeing other students' solutions, and getting a sense of how a subject is taught elsewhere.

Roles for a remote expert

The use of remote experts is seen to have several functions:

- as staff development: "Next time I feel I could do the marking myself"
- as resource: "The ATOM fell at a time when I was very busy, so having X do the marking was very useful"
- to give the primary exposition (e.g. a lecture delivered by video link), though this was not the focus of MANTCHI.
- to give students access to the expert, as postgraduates might have at workshops and conferences, particularly to respond to student questions and suggested solutions to exercises. There is some tendency at undergraduate level, though, to perceive the use of a remote domain expert as an admission of inadequacy by the local deliverer: "Why can't you teach us this yourself?"

This latter issue in fact relates to which view prevails (on the part of learners, teachers, and indeed institutions) of what higher education should be, as described by Perry (1968). This view, which varied within this project, also varies generally and widely with the age of the learner, the institution, but perhaps above all with the discipline. Even at the age of 16, history teaching may be organised around weekly essays based on library reading with the teacher's role one of discussing student work, whereas even at Masters level, you may find chemistry being taught by lecture with the student's role being to reproduce what the teacher says without criticism. Perry's view was that a university's duty is to move students from the latter simplistic view and dependent role to the former. (Nowadays this might be redescribed in terms of acquiring learning or critical thinking skills.) Students however often resist this, and criticising their local teachers for not apparently "knowing" the material is consistent with what Perry regards as the primitive state of seeing teaching as telling, learning as listening, and not having to trouble to examine evidence and alternative views in order to formulate and defend a reasoned view of one's own.

The role of a remote expert will depend upon where on the Perry spectrum a course is pitched. It is however well suited to an approach where students are expected to be able to learn by reading and by attempting to do exercises, but will benefit from (expert) tutors' responses to their questions and examples not directly covered in the primary material. From the teachers' viewpoint, the interviews indicated that the most important function was to deal with those unexpected questions and to comment on student solutions to exercises (which also requires the ability to judge new variations).

It might be nice for local deliverers if they had a remote expert to give lectures, do all the marking, and so on; but the most valuable actions are probably to give some discussion and to answer questions after students have done the reading or heard a lecture, and to give comments on the good and bad features of student solutions (even if they do not decide numerical marks).

Conclusions about future cost benefit studies

Clearly this was an exploratory study, one of whose main contributions is to suggest how to design improved studies in future. The accuracy of the time measures could be improved firstly by having time diaries (logs) kept throughout the period, rather than relying on retrospective recollections. Secondly, some direct observation of the work could be used as a check on the accuracy of the diaries. Thirdly, they would be improved by drawing up and communicating definitions of what time was to be counted. The categories identified in this paper would be the starting point for this.

Having said that, any future study should still continue to look for, and expect to find, new categories of time and other costs and benefits. Future studies should repeat and extend the interview approach of this study in order to do this. Furthermore comparative studies of other teaching situations would be illuminating, as little is known of how higher education teaching work breaks down into component activities.

Finally, the point of cost benefit studies is to support, justify, and explain decision-making: in this case, whether it is rational to join an ATOM scheme for collaborative teaching. The actual measured costs and benefits are just a means to that end. It would therefore be valuable to do some direct studies of such decision-making, for instance by interview or even thinkaloud protocols of teachers making such decisions. That might well bring up more essential categories that need to be studied.

Conclusion: Overall comments on the cost-benefit relationships

The fundamental potential advantages here are that teachers volunteer to author a topic they are already expert in and in exchange adopt an exercise in a topic they think is important, but do not feel expert in; and that material can be re-used in several institutions.

From the authoring viewpoint of creating exercises and the associated handouts (assuming authors use their own material), the author receives materials they do not have to write in exchange for improving something they have or would have already written. The advantage shows up in clear time savings (at least 3 for 1 for a teacher who authors one ATOM and adopts two), in lower subjective effort, and in higher quality topics (i.e. a gain in curriculum quality).

A further potential gain comes from the fact that each exercise will be re-used more often, because it is used at several institutions, than it normally would be. This reduces the authoring cost per delivery, and will often lead to higher quality as feedback leads both to revisions and to the use of stored past solutions and feedback (a feature of MANTCHI not dealt with in this paper).

From the viewpoint of the local deliverer as a course organiser, adopting an ATOM is less work than creating one's own. It is less stressful because its quality is supported by an expert author, and by having been trialled elsewhere, and because its delivery may be supported by a remote expert. Above all, it gives a higher curriculum quality. Within the project, teachers dropped the topics they least valued on their own courses and requested or selected ATOMs from others that they felt most increased the value of the set of topics on their course.

From the viewpoint of the work of local delivery itself, there are three cases. In ATOMs without any use of a remote site, the work is the same. If a remote expert is used, then local deliverers donate some tutor time on a subject they are highly confident about in return for the same time received on a topic they have low confidence about. In contact time, this may not be a saving as some local deliverers will attend as facilitators for the occasion. However such "contact time" does not require the preparation it normally would. For marking, there is a negotiated balance, so no time should be lost. In ATOMs with remote student interaction, there is an extra cost of coordinating the courses at two institutions, which has to be balanced against the pedagogical gains of this form of peer interaction for students, and any relevant gains due to practise with the technology involved.

Thus in return for savings in authoring time, and a gain in curriculum quality (better set of topics covered) and also the quality of individual materials (more often tested and improved, accumulation of past student work as examples), there is either no penalty in net delivery time, or some increase in time spent facilitating (as opposed to more active interaction), with the added staff development reward of the deliverer becoming more confident of their expertise in delivering this material in future without remote support.

Part E:
Conclusion: the relationship with the project as a whole

The project concerned remote collaborative tutorial teaching. It was based on an underlying theoretical idea about tertiary materials, and an underlying technical facility (the internet, specifically the MANs). The development and assessment of these underlying features will be reported on elsewhere. The main evaluation activity was specified in the project proposal as project objective 2 (of 4) "To measure the educational effectiveness of this novel delivery of tutorial support using the method of Integrative Evaluation", and reported on in part B above. The main product of integrative evaluation is formative evaluation recommendations about how to improve overall delivery of teaching (not just how to improve the software or materials themselves). Many of these are collected and presented in part C above. However because integrative evaluation involves extensive study of the learning activities, it allows us to draw some summative conclusions. The evidence on learning outcomes and quality, detailed in part B, suggests that our ATOM materials are of at least as high a quality as other material, and are often though by no means always preferred.

Our studies suggest however that the main gains are in improved curriculum content (replacing less valued topics by ones judged to make for a better course content), and in staff development (expanding the range of topics a teacher is confident of delivering). In addition, the cost benefit analysis (part D above) suggests that there are clear net gains in teaching collaboration organised on our model.

The results of the MANTCHI project, supported by our evaluation studies, might be listed as:

- The invention of two types of material relating to tutorial teaching: ATOMs and TRAILs; and extensive testing of them suggesting that they at least maintain and perhaps raise learning quality. Reasons for improved quality include a) each item is written by an expert on that topic (not just on the subject area as a whole), b) the material is delivered more often, and is likely to be improved after each iteration, c) the inclusion where possible of "TRAILs" (example questions, solutions, and tutor feedback on those solutions) improves the supply of information for learners of a kind that is often scarce because it is cost-intensive to provide only by the usual method of personal marking of a student's own assignments.
- A method of true reciprocal collaborative teaching, that is rare and perhaps unique; and a cost benefit analysis suggesting clear net gains from adopting it.
- The chief gain seems likely to be in curriculum quality. While the educational quality of each item (given a fixed topic) seems to be at least adequate and probably superior in this format, the selection of topics on a course when these are available is improved in the academic judgement of the course organisers.
- A further gain noted by some participants is in effect that of staff development: the adoption of materials together with active support can lead later to staff feeling able to deliver these materials without that support in future.

Appendix 1
List of evaluation studies

Evaluation of HCI Tutorials/Courses/Collaborations

1) HCI Module, B.Sc. Level 4, Napier University. David Benyon
Evaluation of the teaching of Entity-Relationship Modelling of Information Artefacts (ERMIA).
Observation and discussions with students and staff. (Feb.-March 1997)
Report: 12/5/97 by Margaret Brown (submitted to David Benyon and circulated to MANTCHI and external assessor)

2) CSCW Module, BSc Computer Studies & BSc Computer Information Systems;
Level 4, Glasgow Caledonian University. Julian Newman
Observation of a seminar on 19/3/97; informal discussion with students about the CSCW course; use of a Resource Questionnaire. The lecturer's concerns included student use or otherwise of academic papers when researching the topic for their seminar. (March 1997).
Report 26/3/97 (additional note: 12/5/97) by Margaret Brown (submitted to Julian Newman and circulated to MANTCHI and external assessor)

3) Interactive System Design Course: B.Sc. Level 4, B.Eng. Level 4, and M.Sc.
Heriot Watt University. Alistair Kilgour
Evaluation of a "tutorial" on Dialogue Description Formalisms (Statecharts): Pre and Post Tutorial Questionnaires administered by Prof. Kilgour. (Feb. 1997)
Report 16/4/97 (additional note 12/5/97) by Margaret Brown (submitted to Alistair Kilgour and circulated to MANTCHI and external assessor)

4) Interactive System Design Course: B.Sc. Level 4, B.Eng. Level 4, and M.Sc.
Heriot Watt University. Alistair Kilgour
Evaluation (at a distance) of the Interactive System Design Course using a Resource Questionnaire. (May 1997)
Report 8/10/97 (updated from 22/8/97) by Margaret Brown (submitted to Alistair Kilgour)

5) Collaboration: GIST Day: 15/5/97
Evaluation of the Collaboration between Glasgow University and Dundee University;
the Presentations by Staff and Students from GU and Dundee; the Video Recording of these Presentations by MANTCHI.(May 1997)
Report: 24/9/97 by Margaret Brown, Mike Kavanagh and Norman Gray (submitted to Phil Gray, Steve Draper, Alistair Kilgour)

6) Conversion M.Sc. : M.Sc. Information Systems/Software Technology: HCI Module: Napier University.
Alison Varey & Alison Crerar
Observations of the lecturers' assessments of students' prototypes of "walk-up and use" systems. (May 1997)
Report: 23/6/97 (submitted to Alison Varey)

7) M.Sc. IT Course: HCI Module. Glasgow University. Steve Draper
Evaluation of an HCI Course using observation of lectures and labs; discussions with students;
Confidence Logs and a Resource Questionnaire. (Jan. - March 1997)
Report (In form of results but no discussion; submitted to Steve, August 1997)

8) B.Sc. Computing/Information Systems: Level 3 HCI Module. Napier University. Alison Varey
Evaluation of Lab and class based tutorials using a Tutorial Provision Questionnaire and observations of one Class-based Tutorial; one Lecture; peer evaluation of interfaces to multimedia information systems (March-May 1997)
Report submitted to Alison Varey, Feb. 1998

9) HCI Module, Level 1 Computing Science, Glasgow University
Evaluation of an HCI Level 1 Course, using observation of tutorials and labs; a Resource Questionnaire; student marks; Questionnaire to tutors. (Jan-June 1997)
Report : not yet completed (to be submitted to Chris Johnson)

Evaluation of MANTCHI ATOMs

10) HCI Module, B.Sc. Level 4, Napier University.
Evaluation of the use of the **Statecharts ATOM** and the **ERMIA ATOM** at Napier University. Using pre and post-tutorial Questionaires; pre and post-assignment. Questionnaires; observation; discussion with students and staff. (Oct. -Nov 1997)
Report (28/11/97) (submitted to David Benyon, Helen Lowe, Sandra Foubister, Steve Draper)

11) HCI Module, B.Sc. Level 4, Glasgow University.
Evaluation of the use of the **Statecharts ATOM.** Using pre-tutorial Questionaires; pre and post-assignment Questionnaires; Resource Questionnaire; observation; discussion with students and staff. (Oct.-Dec. 1997)
Draft Report: submitted to Phil Gray April 1998)

12) CSCW Module, Level 4 Napier and Caledonian Universities.
CSCW ATOM (refer to Evaluation 2)
Evaluation of the use of **desk top video conferencing** between Universities. Observations of presentations on **CSCW** by students of one University to those in another.
Draper.Final Report (10/2/98) to Peter Harper, Julian Newman, Michael Smythe and Steve Draper

13) Interactive System Design Course: B.Sc. Level 4, B.Eng. Level 4, and M.Sc.
Heriot Watt University. Alistair Kilgour(refer to Eval. 4)
Evaluation of the use of the **UAN and Cognitive Walkthrough ATOM**, the **Statecharts ATOM & the ERMIA ATOM**. Using a combined Resource Questionnaire; Confidence Logs; discussion with students and staff; Post-Course e-mail Questionnaire. (Feb-May 1998)
Basic numerical data (no student comments): submitted to Alistair Kilgour April 1998.
Draft Report submitted to Alistair Kilgour July 1998

14) HCI Module (Graphics & Databases), B.Sc. Level 3 Glasgow University.
Evaluation of the use of the **UAN and Cognitive Walkthrough ATOM**.
Using a Resource Questionnaire; Confidence Logs;observation; discussion with students and staff. (Jan- March 1998)
Basic numerical data (no student comments): submitted to Phil Gray (March 1998)
Draft Report submitted to Phil Gray (May1998)

15) M.Sc. HCI; M.Sc. DMIS: Computers in Teaching & Learning Module, Heriot Watt University. Alison Cawsey & Patrick McAndrew.
ATOM on Evaluating CBL: Evaluation of the ATOM by observing 2 tutorials run by Steve Draper using video conferencing; discussions with students and staff; a Resource Questionnaire & Confidence Log. (Feb.—March 1998)
Report of first video conference submitted to Patrick McAndrew, Alison Cawsey and Steve Draper (12/2/98)
Basic numerical data (no student comments): submitted to Alison Cawsey, Patrick McAndrew & Steve Draper (March 1998)
Draft Report submitted: June 1998

16) M.Sc. IT Course: HCI Module. Glasgow University. Steve Draper (refer to Eval 7)
Evaluation of the development and use of **student generated web-based collaborative lecture notes** for a HCI Course using observation of lectures and labs; discussions with students; a Resource Questionnaire; Confidence Log. Post-exam Questionnaire (Jan. - May 1998)
Data from Resource Questionnaire not yet compiled and analysed.
Basic Results from post-exam questionnaire submitted to Steve Draper (May 1998)

17) M.Sc. OOSE: HCI Module. Napier University. David Benyon
Evaluation of the use of 5 ATOMs: **UAN & Cognitive Walkthrough; Statecharts; ERMIA; Web site Design; Evaluating Web sites.**
Proposed Evaluation: Short e-mail questionnaires on each ATOM; Resource Questionnaire on whole course; discussions with students and staff. (Mar. -May 1998)
Actual evaluation: e-mail Resource Questionnaire on UAN & Cognitive Walkthrough ATOM: end of course discussion on all 5 ATOMs; end of course Resource Questionnaire on all 5 ATOMs (paper/e-mail).
Basic Results submitted to David Benyon June 1998, updated July 1998.

18) **BSCW** Collaboration Glasgow Caledonian University (M.Sc) and Heriot-Watt University (M.Sc.) (Feb-March 1998) Cancelled (refer to Evaluation 4)

19) Building Interactive Information Systems Module, MSc in Advanced Information Systems, Glasgow University. Phil Gray.
Evaluation of the use of the **ERMIA ATOM** by 6 students, December 1997. Informal discussions with the students.
Report: 18/12/97 to Phil Gray.

20) Conversion M.Sc. : M.Sc. Information Systems/Software Technology:
HCI Module: Napier University. Alison Varey & Alison Crerer.
Formative Evaluation ATOM (refer to Evaluation 6)
In conjunction with Level 4, Glasgow Caledonian University. Helen Lowe
Observations of the lecturers' assessments of students' prototypes of "walk-up and use" systems. (May 1998) some of which had been evaluated by remote evaluators.
Report: submitted to Alison Varey July 1998
In addition, help was given to the lecturer, Alison Varey, to develop a Resource Questionnaire to evaluate the module and specifically the "remote" evaluations.

Additional Work

September-October 1997: Discussions with M.Sc. IT students at Glasgow University on their thoughts on feedback, discussions via the MAN etc.
Report (6/10/97) (submitted to Jane Reid, Steve Draper)

Appendix 2
MANTCHI Resource Questionnaire:
UAN & Cognitive Walkthrough ATOM
Heriot -Watt University:MSc and Level 4

Date: 9/3/98 Gender: M ❑ F ❑ Age: .....Matric. Number:............

Degree: M.Sc. ❑ B.Sc. ❑ B.Eng. C&E ❑ B.Eng.ISE ❑

The following questionnaire is concerned mainly with your use of the **UAN & Cognitive Walkthrough ATOM** and other learning resources during your course. All answers are confidential and will not be attributed to individual students. Your matric. number is only required in order to link this questionnaire to any future questionnaires. Thank you for your participation in this project.

Dr. Margaret I. Brown, Dept of Psychology, Adam Smith Building, University of Glasgow, Glasgow, G12 8QQ.
mag@psy.gla.ac.uk MANTCHI: http://greenock-bank.clyde.net.uk/mantchi/

Q.1. Which **Formalisms** have you already studied on this or previous Formal Methods Courses or in your own time? (Only tick UAN and Statecharts if you studied them before 1998)

UAN❑ ERMIA❑ Statecharts❑ BNF❑ State Transition Diagrams❑ Petri Nets ❑ Event CSP❑

Object Z❑ Other ❑ (Please name)

Q.2. How easy do you find it to understand the underlying concepts behind Formalisms and their use and application ? Please indicate on the scale of **0—4** below.

Extremely easy **0 1 2 3 4** *Extremely difficult*

Explain:

Q.3. Please indicate by ticking the relevant box how confident you feel that you are able to complete the following objectives.

Objective	No Confidence Whatsoever	Little Confidence	Some Confidence	Confident	Very Confident
write a UAN task description of a small-scale computer-based task					
using this task embedded in a scenario, perform a cognitive walkthrough					
based on the cognitive walkthrough, identify simple potential usability problems with the task.					

a. Comment on any effects the Group work (including discussions with other Groups had on your reported confidence levels for the objectives**.**

b. Comment on the use of Group work as opposed to individual work in your University courses. (e.g. benefits and disadvantages: indicate group size referred to).

Q.4a. Learning Resources/Activities. In the table below, mark each resource/activity you used while learning about **UAN & Cognitive Walkthrough**. If used, please mark how useful you consider each resource or activity was to you in learning and understanding UAN & Cognitive Walkthrough and also its use and application. Please give reasons for your answers. (Remote Expert = Phil Gray: GU)

Activity/ Resource	**Used?**	***If used / attended* Usefulness of Activity / Resource**					**Reason for answer**
	tick if used	not at all useful	not very useful	useful	very useful	extremely useful	
1.Introduction to UAN& Cognitive Walkthrough by Prof. Kilgour							
2. Textbook (Hix & Hartson)							
3. Handout from Prof. Kilgour (Hartson's notes)							
4. Web resources on UAN written by Remote Expert							
5. Reference on ATOM Page (Univ. of Toronto)							
6. Group Work on UAN & Cognitive Walkthrough ATOM tasks							
7. Receiving feedback etc from the Remote Expert on your Group's solutions to the tasks							
8. Access to the solutions submitted by other Groups (from HW)							
9. Access to feedback from the Remote Expert on other Groups' solutions to the tasks							
10. Comparing own Group's solutions and feedback with those from other Groups							
11. Discussion using ATOM discussion forum (Hypernews)							
12. Asking questions to Remote Expert (e-mail, discussion forum etc.)							
13. Discussions with Prof. Kilgour							
14. Discussions with other students on course							
15. Re-accessing the submissions and feedback while preparing your Group Reports							
16. Other resource (please specify)							

Q.4b. If you considered any of the resources "not very" or "not at all useful" what did you do, or what will you do to compensate?

Q.4c. Will you look at the submissions from Glasgow University students when they become available as an additional ATOM Resource?

YES ❑ **NO** ❑

How useful do you consider they will be as an additional learning resource?

not at all useful *not very useful* *useful* *very useful* *extremely useful*

Comment

Q.5. Time available to use resources or perform activities for the **UAN & Cognitive Walkthrough ATOM**

Please list the resources that you consider you had:

a) not enough time to use b) too much time allocated for their use

Refer to the resources listed in Q.4. You can state the resource number instead of the name of the resource.

Resources/Activities: **insufficient time for use:**	**Resources/Activities:** **allocated too much time:**

Q.6. Did you experience any difficulty gaining access to any resources / activities during the use of the **UAN & Cognitive Walkthrough ATOM**?

YES ❑ **NO** ❑

Please list the resources/activities explain the problems. *(Refer to the resources listed in Q.4. and also to the labs, computers, software, technical help etc. You can state the resource number instead of the name of the resource.)*

Q.7. Did you have any problems when doing the following:

Activity	**tick if you had a problem**	**Explain any problems**
Understanding what you were expected to do for the assignment		
Completing the UAN & Cognitive Walkthrough tasks		
Creating the Web page for your solution		

Q.8 **a.** Professor Kilgour provided paper-based copies of the ATOM Resources (except Toronto nformation)

Had you already printed the ATOM information and scenario from the Web?

YES ❑ **NO** ❑

b. Did you use the information/reference papers in paper-based and/or Web-based form?

Paper-based form ❑ **Web-based form** ❑

Explain:

c. How and where do you download and print information from the Web?

Q.9. **a.** Was the "level" of the material in the ATOM right for you?
Indicate on the scale of 0 — 4 below.
far too easy 0 1 2 3 4 far too difficult

b. Was the "workload" in the ATOM right for you? *Indicate on the scale of 0 — 4 below.*
far too little 0 1 2 3 4 far too much

Comment:

Q.10 **a.** How much did you benefit by taking part in an exercise with input from a remote expert **(The UAN & Cognitive Walkthrough ATOM)**?
Indicate on the scale of 0 — 4 below .

not at all 0 1 2 3 4 substantially

b. Please list the benefits and any disadvantages of the collaboration.

benefits	**disadvantages**

c. Were any benefits or disadvantages unexpected? **YES** ❑ **NO** ❑
Mark them with a tick **(√) on your lists in Q.10b.**

Q.11. What do you consider contributed most to your understanding of the underlying concept of **UAN & Cognitive Walkthrough?**

Q.12. a. Do you consider that you still require extra information/training to help you learn/understand/apply **UAN & Cognitive Walkthrough**?

YES ❑ **NO** ❑

b. Explain the type of extra information/training you consider you require.

c. What else would have helped you to complete the ATOM tasks?

Q.13. You have started studying **ERMIA** by accessing and using the MANTCHI ERMIA ATOM. Have you any comments on your use of this ATOM so far?

Q.14. Please give any other comments on your use of MANTCHI ATOMs and the use of other Web-based ATOMs and on-line discussion forums on University Courses in the future.

Q.15. Have you any other comments?

References

Brown, M.I., Doughty,G.F., Draper, S.W., Henderson, F.P. and McAteer, E. (1996) "Measuring Learning Resource Use." Computers and Education vol.27, pp. 103-113.

Doughty, G. (1996a) "Technology in Teaching and Learning: Some Senior Management Issues", "Deciding to invest in IT for teaching" in TLTSN Case Studies (HEFCE)

Doughty, G. (1996b) "Making investment decisions for technology in teaching" (University of Glasgow TLTSN Centre) [WWW document] URL http://www.elec.gla.ac.uk/TLTSN/invest.html

Doughty,G., Arnold,S., Barr,N., Brown,M.I., Creanor,L., Donnelly,P.J., Draper,S.W., Duffy,C., Durndell,H., Harrison,M., Henderson,F.P., Jessop,A., McAteer,E., Milner,M., Neil,D.M., Pflicke,T., Pollock,M., Primrose,C., Richard,S., Sclater,N., Shaw,R., Tickner,S., Turner,I., van der Zwan,R. & Watt,H.D. (1995) Using learning technologies: interim conclusions from the TILT project TILT project report no.3, Robert Clark Centre, University of Glasgow ISBN 085261 473 X

Doughty,P.L. (1979) "Cost-effectiveness analysis tradeoffs and pitfalls for planning and evaluating instructional programs" J. instructional development vol.2 no.4 pp.17,23-25

Draper, S.W. (1997a, 18 April) "Adding (negotiated) management to models of learning and teaching" Itforum (email list: invited paper) [also WWW document]. URL: http://www.psy.gla.ac.uk/~steve/TLP.management.html

Draper, S.W. (1997b) "Prospects for summative evaluation of CAL in higher education" ALT-J (Association of learning technology journal) vol.5, no.1 pp.33-39

Draper, S.W. (1998) CSCLN ATOM [WWW document]. URL http://www.psy.gla.ac.uk/~steve/HCI/cscln/overview.html

Draper, S.W., & Brown, M.I. (1998) "Evaluating remote collaborative tutorial teaching in MANTCHI" [WWW document] URL http://www.psy.gla.ac.uk/~steve/mant/mantchiEval.html

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 vol.26 no.1-3, pp.17-32

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

Mayes, J.T. (1995) "Learning Technology and Groundhog Day" in W.Strang, V.B.Simpson & D.Slater (eds.) Hypermedia at Work: Practice and Theory in Higher Education (University of Kent Press: Canterbury)

MANTCHI (1998) MANTCHI project pages [WWW document] URL http://mantchi.use-of-mans.ac.uk/

Perry, W.G. (1968/70) Forms of intellectual and ethical development in the college years (New York: Holt, Rhinehart and Winston)

Reeves,T.C. (1990). "Redirecting evaluation of interactive video: The case for complexity" Studies in Educational Evaluation, vol.16, 115-131.

Reigeluth,C.M. (1983) "Instructional design: What is it and why is it?" ch.1 pp.3-36 in C.M.Reigeluth (ed.) Instructional-design theories and models: An overview of their current status (Erlbaum: Hillsdale, NJ)

TILT (1996) TILT project pages [WWW document] URL http://www.elec.gla.ac.uk/TILT/TILT.html (visited 1998, Aug, 4)