Evaluating remote collaborative tutorial teaching in MANTCHI

Stephen W. Draper & Margaret I. Brown
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

Abstract

This chapter overviews the integrative evaluation studies done during the MANTCHI project. About 20 studies were done of the many deliveries. In this chapter, some main conclusions each drawing on evidence from multiple studies are presented, together with discussion on the strengths and weaknesses of the method employed. Overall, the evaluation studies provided both formative and summative results. The formative information for improving the overall delivery of the material is exemplified by a selection of issues and recommendations, and a sample of the evidence on which they were based. The summative evidence on learning outcomes and quality suggests that the project materials are at least as effective as other material, although by no means always preferred by students. 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.

Introduction

This chapter discusses the evaluation work on the MANTCHI project (described below) as a whole. It concerned collaborative 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. The project focussed on "tutorial" material rather than primary exposition like lectures, and was interested in the re-use of "tertiary" material such as past student solutions, and whether this was useful to students.

The evaluation work comprised about 20 studies. Full reporting of so many studies, even though of modest size, is not possible within a single chapter. On the other hand, such a chapter can (and this one 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 chapter is divided into three parts, the first dealing with the overall approach, the second with particular findings, and the third reviewing the adequacy of the evaluation methods. 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. The third part discusses the successful and unsuccessful aspects of our approach, and offers suggestions for future improvements to the method.

Part A: 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). This material might for instance consist of an exercise created at another site and backed up by a remote expert (often the author), who might 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 chapter, 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" (Tertiary Reusable ATOM Instantiated for Learning).

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 chapter offers a general report on that evaluation activity.

The evaluation approach

The evaluation method 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) is 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 the appendix.

Part B: 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 students 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 issue of learning effectiveness we designed comparable measures for this, whereas most of the management issues emerged from open-ended measures (and complaints). However this 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: 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 students. It may be another manifestation of the importance of active learner manipulation of knowledge representations discussed by Ravenscroft et al. (1998).

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 three 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 good 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 some 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 have 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: "[TRAILs] 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. Here we present a sample of such findings, together with suggestions for responses as they were circulated within the project. (A complete set is available in Brown and Draper, 1998.) 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. For the first item, we give details of the evidence on which the finding and recommendations were based; for other items (for reasons of space) details of the evidence are omitted. Should it be important to clarify an issue first identified by open-ended measures, then a more systematic measure could 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.

Sample issues

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 superceded 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 on which they can make notes. 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.

An example of the evidence (Web-based vs. Paper Resources)

The evidence on which the conclusion and recommendation on web-based vs. paper resources were based was as follows. 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!"

Content of ATOM

Students would find it useful if the ATOM contained 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 approximate date on which feedback will become available.

Delivery/timing

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.)

Students should be alerted when feedback on their solutions is available. They should also be alerted when feedback on the solutions from other universities is available. That is, posting them on the web without emailing an announcement is unsatisfactory.

Time

Instructions about the ATOM resources and assignments have to be sent to students in plenty of time. 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.

Domain 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 of whether the lecturer will be "facilitating" and not teaching and that the subject will involve "resource-based" learning utilising the ATOM Web-based resources and a domain expert.

Collaboration between Students from different Universities

One of the ATOMs involved students at two Universities. Comments from students at both universities indicated some rivalry and comparison was also used in the feedback. Although this can be a good thing, we have to be careful to avoid the collaborations discouraging some students from actively participating.

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 we should be aware of and should discuss. Students being required to write a report involving the topics studied on two ATOMs appeared to be successful at one University.

Conclusion

Many of these points will 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 or never 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 C: Review of the evaluation method

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 our 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 is reported in another chapter (Draper and Foubister, 1998). 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.

Overall 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?

Acknowledgements

The authors acknowledge the support of Scottish Higher Education Funding Council (SHEFC) through the Use of MANs Initiative for the MANTCHI project. Thanks to all the people who contributed to MANTCHI (see the web site for a list of personnel: MANTCHI, 1998) without whom there would have been nothing attempted, and to all the anonymous students without whom there would have been no learning and no evaluation data.

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 it's 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 information)

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.

Brown,M.I. and Draper, S.W. (1998) Lessons on delivering tutorial teaching and ATOMs based on MANTCHI evaluation studies Technical report [WWW document]. URL http://www.psy.gla.ac.uk/~steve/mant/lessons.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

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., 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

Draper, S.W. and Foubister,S.P. (1998) "A cost-benefit analysis of remote collaborative tutorial teaching" This volume

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/

Ravenscroft,A., Tait,K. & Hughes,I. (1998) "Beyond the media: Knowledge level interaction and guided integration for CBL systems" Computers and Education vol.30 no.1-2, pp.49-56

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)