Sharples, M. (2009) Methods for Evaluating Mobile Learning. In G.N. Vavoula, N. Pachler, and A. Kukulska-Hulme (eds), Researching Mobile Learning: Frameworks, Tools and Research Designs. Oxford: Peter Lang Publishing Group, pp. 17-39.

Methods for Evaluating Mobile Learning

Mike Sharples
Learning Sciences Research Institute, University of Nottingham

Overview
Mobile learning differs from learning in the classroom or on a desktop computer in its support for education across contexts and life transitions. This poses substantial problems for evaluation, if the context is not fixed and if the activity can span formal and informal settings. There may be no fixed point to locate an observer, the learning may spread across locations and times, there may be no prescribed curriculum or lesson plan, the learning activity may involve a variety of personal, institutional and public technologies, it may be interleaved with other activities, and there may be ethical issues concerned with monitoring activity outside the classroom. The chapter indicates issues related to evaluation for usability, effectiveness and satisfaction and illustrates these with case studies of evaluation for three major mobile learning projects. The Mobile Learning Organiser project used diary and interview methods to investigate students' appropriation of mobile technology over a year. The MyArtSpace project developed a multi-level analysis of a service to support learning on school museum visits. The PI project has employed critical incident analysis to reveal breakthroughs and breakdowns in the use of mobile technology for inquiry science learning. It is also addressing the particular ethical problems of collecting data in the home.

Introduction
Mobile learning is not simply a variant of e-learning enacted with portable devices, nor an extension of classroom learning into less formal settings. Recent research has focused on how mobile learning creates new contexts for learning through interactions between people, technologies and settings, and on learning within an increasingly mobile society (Sharples, Taylor, & Vavoula, 2007). For example, a young child visiting a science discovery centre may create a favourable context for learning, through interactions with the exhibits, a multimedia guide, conversation with parents, and observations of other visitors. In a mobile society, people are continually creating such opportunities for learning though a combination of conversation on mobile devices, context-based search and retrieval of information, and exploration of real and virtual worlds.

New modes of learning are being designed, including mobile game-based learning (Schwabe & Goth, 2005), learning from interactive location-based guides (Damala & Lecoq, 2005; Naismith, Sharples, & Ting, 2005), and ambient learning (Rogers et al., 2004). While these offer opportunities to support personalisation and to connect learning across contexts and life transitions, they also pose problems in evaluating learning processes and outcomes. If mobile learning can occur anywhere, then how can we track and record the learning processes? If the learning is interwoven with other everyday activities, then how can we tell when it occurs? If the learning is self-determined and self-organised then how can we measure learning outcomes? These are difficult questions, with no simple answers, yet it is essential to address them if we are to provide evidence of the effectiveness of mobile learning.

With a few notable exceptions (see e.g. Valdivia & Nussbaum, 2007) most studies of mobile learning have either provided evaluations in the form of attitude surveys and

interviews ("they say they enjoy it"), or observations ("they look as if they are learning") (Traxler & Kukulsa-Hulme, 2005). Although surveys, interviews and observations can illuminate the learning process, they do not provide detailed evidence as to the nature and permanence of the learning that has occurred.

Issues in evaluating mobile learning

To propose appropriate evaluation methods for mobile learning, we need to understand what distinguishes mobile learning from classroom learning or learning with desktop computers. Below are some distinctive aspects of mobile learning. Not all are definitive, for example informal learning may be carried out with fixed computers, but taken together they indicate the space of learning activities for which evaluation methods need to be applied. A useful framework to map this space is the distinction (adapted from Livingstone, 2001) between whether the learning is initiated by the learner, or externally (e.g. a teacher or a curriculum) and whether the learning process is managed by the learner or others, see Table1.

	External management	*Learner management*
External initiation	Formal learning	Resource-based learning
Learner initiation	Voluntary learning	Informal learning

Table 1 Initiation and management of learning (adapted from Livingstone, 2001)

Mobile learning may be mobile (but not necessarily)

If we take as a definition of mobile learning "*Learning that happens across locations, or that takes advantage of learning opportunities offered by portable technologies*."[1] then it may occur either on the move, or in a fixed location such as a classroom where the learners are using portable devices. Existing methods for evaluating curriculum-led learning in the classroom (such as assessments of curriculum learning gains and classroom observation studies) can equally be applied to learning with handheld devices, and indeed already are so, since classroom maths learning typically involves the use of pocket calculators. What is novel, is the evaluation of technology-enabled learning on the move, with the learner travelling while learning, or learning that spans locations and times. This is not easy to capture as there may be no fixed point in time or space to locate an observer or a video camera. To complicate matters further, mobile learning may be distributed, involving multiple participants, spread over many locations, some or all on the move.

Mobile learning may occur in non-formal settings

Many studies of mobile learning have been in non-formal educational settings, such as museums and field trips. Such settings are designed to support activities that may be initiated by parents or schools, but are managed by the learner Thus, we may know what variety of learning opportunities are on offer, but not how the students will engage in the activity, making choices and creating a path through the exhibits and

[1] A definition contributed to the Wikipedia Mobile Learning page by the author, and as yet uncontested (adapted from O'Malley et al., 2005).

resources. Contrast this with a traditional classroom, where the lesson plan provides an indication of what should be learned, at what time.

Mobile learning may be extended and interleaved

An appealing aspect of mobile learning is that it can support the learner over a long period of time, for example in learning a foreign language, and that it can be interleaved with other activities, such as taking a plane journey, or being a tourist. A consequence is that it may not be possible to determine when the learning begins and ends, nor when a person is deliberately learning or just enjoying an activity (which itself may lead to unintended learning).

Mobile learning may involve a variety of personal and institutional technologies

During a typical day a university student may move around the campus engaging with a variety of technologies and resources that can support learning such as desktop computers, personal laptop computer, multimedia lecture rooms, electronic dictionary, MP3 player, mobile phone, and a variety of books and notes. A tourist may learn from a personal phone or MP3 player, an audio guide, a multimedia booth, a printed guidebook and a human guide. To evaluate the effect of one such device, such as a mobile phone, in this melange of technology would either require setting up an artificial experiment (for example, requiring the student to learn everything though the phone) or to try and isolate the specific learning effects of individual devices.

Mobile learning presents particular ethical problems

Although it may be technically possible to monitor learning activities outside a formal setting such as a classroom, for example by logging all use of students' laptop computers or setting up cameras in museums or tourist sites, there may be ethical objections to doing so. These include getting permission from all participants to be monitored for research purposes and allowing participants the right to choose when to be monitored. This is particularly sensitive when the participants are children, or the learning is part of an assessed curriculum, since those involved may feel coerced into participating. An evaluation must address both the specific aspects of ethical research and broader issues of the rights of children, at different ages, to escape from continual monitoring and to be free to play and explore without continual pressure to learn.

What do you want to know?

It may be obvious, but nevertheless needs to be stated, that the appropriate method of evaluation depends on what the evaluators want to know. Choice of method also depends on who needs to know the results and how they will be used. Evaluation as part of education research will be concerned with understanding how fundamental processes of learning can be mediated, enhanced and transformed. Evaluation to inform design will focus on intervention and enhancement, examining how a combination of technologies and activities can best be developed to address problems and provide new learning opportunities. Evaluation for policy makers needs to provide evidence of learning gains or changes, either through comparison with existing approaches, or by showing how mobile learning can create radically new opportunities, such as linking people in real and virtual worlds. A useful way to approach the evaluation, for any stakeholder, is to address usability (will it work?), effectiveness (is it enhancing learning?) and satisfaction (is it liked?).

Usability

If the aim is to improve the technology, then there are well-established methods of usability testing in the lab, such as heuristic evaluation (Molich & Nielsen, 1990), that have been successfully transferred to mobile devices. Evaluating usability in the field is more difficult, but the technology itself can assist. The software may be programmed to log user interactions and their times, to show a timeline of user activity or to replay the interactions. One important point is that the technology may not be able to log its own breakdowns, nor what happens when it is deliberately switched off by the user, but it can show when these start and end. A more innovative possibility is to use the inbuilt multimedia capabilities in devices such as mobile phones and laptop computers to audio record and photograph activities, continually, or at timed intervals, or random intervals, or when initiated by the user. For examples of the use of the mobile technology itself to gather evaluation data see also Wali et al., Hooft, and Trinder et al. in this volume.

Effectiveness

The effectiveness of mobile learning depends on the educational aims and context. What is useful for school or college learning may have little relevance to the learner's informal learning. Conversely, what a person learns outside the classroom may not match the immediate aims of the curriculum, though it may be valuable in supporting aspects of lifelong learning such as carrying out independent research or engaging in social interaction. Thus, any assessment of the effects of mobile learning must be related to the context of the activity and its intended aims. Is the aim to learn a topic, to develop specific skills, or to support incidental and lifelong learning? Is it initiated and managed by the learner, or externally?

Satisfaction

Satisfaction with mobile learning is superficially the easiest to assess, through attitude surveys or interviews with learners, so it is no surprise that many research papers report these as the main or only method of evaluation. Typically, a paper will present a mean response of, say, 3.8 on a 5 point Likert scale to the question "Did you enjoy the experience" as if this were evidence of unusual satisfaction, or even of productive learning. It is neither. Yet almost all results of attitude to novel technology lie within the range of 3.5 to 4.5, on a 5 point scale for satisfaction, regardless of technology or context. More specific questions, such as "Was the technology easy to use?" merely provoke further questions, such as "By comparison to what?"; "For what tasks?". A more refined method of assessing satisfaction is through product reaction cards. The Microsoft Desirability toolkit (Benedek & Miner, 2002) includes 118 cards with words such as 'confusing', 'flexible', 'organised', 'time-consuming' that can be used to assess reaction to a technology or to an experience. Typically, people are asked to select the cards that best relate to their experience and these could form the basis of an interview (e.g. "what did you find confusing about your experience?").

Case studies

The remainder of the chapter offers three case studies in evaluation of mobile learning. The projects have been chosen because they present particular difficulties, in assessing learning over time, or in different contexts, or because of ethical issues. The evaluation methods are not intended as definitive solutions, but in the spirit of object lessons to be examined critically, to gain insight into the successes and limitations of particular evaluation methods. The aim here is not to report results of the projects, but rather to describe the methods of evaluation and to indicate how

successful these were in assessing learning processes and outcomes and in revealing usability, effectiveness and satisfaction.

Mobile Learning Organiser

The student learning organiser project (Corlett, Sharples, Bull, & Chan, 2005) was an early attempt to provide university students with a personal device to support their learning over a long period of time. The motivation for the project was that many business people carry personal organisers offering a set of tools – calendar, contacts list, email, to do list, etc – to manage their working lives. These were designed to support office work rather than learning, so is there a value in developing an analogous 'Mobile Learning Organiser' to assist university students in managing their studies?

Seventeen students on an MSc course were loaned iPAQ Pocket PC devices, with wireless LAN connection but no phone, for one academic year. The devices were equipped with a custom-designed Mobile Learning Organiser that included a Time Manager (for viewing course timetables and lecture slots), a Course Manager (to browse and view teaching material), a Communications Manager (for email and instant messaging when in wireless range) and a Concept Mapper. In addition to these tools, the students could access the full range of Pocket PC software (including calendar, email, instant messaging and files through the normal Pocket PC interface) and were also encouraged to personalise the devices by downloading any media and applications they wished.

Thus, the context of the study was that the students could engage in a variety of activities with the devices, including ones not directly related to learning such as downloading music, in any location within and outside the university, over a period of a year. Given this range and duration, and evaluators' interest in understanding the process of technology adoption and patterns of use, we adopted a mixed-methods approach to evaluation.

The students were asked to complete questionnaires at 1, 4, 16 and 40 weeks after they were issued with the devices. They were asked to indicate the frequency of use of the device ('many times a day', 'at least once a day', at least twice a week', 'less that twice a week') and to rate each provided tool as 'very useful', 'useful', 'possibly useful', 'probably not useful', 'not useful' or 'don't know'. They were also asked to name the tools that made the greatest impact on their learning, personal organisation and entertainment. The freeform answers were collected under generic headings. The questionnaires were successful in revealing changes in use over the year. For example, ratings for the timetable tool increased over the year (59% of the students rated it as 'useful' or 'very useful' after 4 weeks, 64% after 16 weeks, and 84% after 10 months), the instant messaging tool increased in popularity at the end (59%, 50% and 71%) and the perceived value of the course materials decreased over time (59%, 43%, 41%). The concept mapper was least successful, declining from 14% at the start to none indicating it as being useful by the end of the trial.

Taken alone these figures are intriguing, but not particularly revealing. Each survey, however, was followed by a focus group meeting with all the students, to discuss the meaning of the results and also to raise other issues and problems. These meetings helped in interpreting the raw results. For example, one reason for the decline in usefulness of the course materials was that later in the course students were engaged in project work rather than structured learning. They also illuminated general and specific usability problems. Thus, battery life was a major factor in the

decision of some students to abandon their devices, in particular when some students left them behind over the Christmas vacation and the battery discharged losing their data.

The students were also asked to complete written logbooks of their daily activities with the PDA devices, including the location, duration and type of activity. The logbooks revealed patterns and frequency of use across locations during the first six weeks of the project. These provided some unexpected interactions between location and activity, for example:

- Although email was synchronised to the device, students only tended to use this when in an area covered by the campus wireless network.
- Participants used the calendar and timetabling in every location as they had need. So for some students, the PDA became a replacement for traditional diaries.
- Some students reported regularly reading course materials, offline web content and e-books when at home or in their dormitories, even though they all had access to a desktop computer at home.

A final survey was administered at the end of the project. The questions addressed specific issues that had arisen from the earlier surveys and focus groups. Students were asked to rate statements on five-point Likert scale from 'Strongly Agree' to 'Strongly Disagree'. The responses were then weighted from 2 to -2. The sum of weighed responses from each question was then used to measure overall agreement/disagreement. Only four mean responses were in the range +/- 0.5 to 2.0:

Having to use the iPAQ hindered my learning	- 0.7
I found battery life a significant problem	+ 1.1
I felt uncomfortable using the iPAQ because I didn't know how to use it	- 0.8
The advantages of having an iPAQ outweighed the drawbacks of taking part in the trial (attending meetings, doing questionnaires etc.).	+ 0.6

Table 2. Survey questions for which the responses were in the range +/- 0.5 to 2.0[2]

The evaluation methods were designed to be interpreted as a whole, to reveal patterns and trends in technology adoption. For a more general discussion of adoption of mobile technology for learning see Waycott (2004). The methods were successful in revealing some clear modes of use. Frequency of use fell over the period, with 60% using the devices less than twice a week by the end of the project compared to 18% at the start, however 22% continued to use the PDA many times a day, a similar percentage to that at the start of the project. The students made considerable use of the calendar and timetabling features as well as the communications tools. Content optimized for the PDA was well used, and there was a request from some students that more resources should be made available in PDA format, including administrative information.

An unexpected result, given the aims of the study, was that there was no conclusive evidence of need for a specifically designed suite of tools in addition to those already

[2] Minus figures indicate a disagreement with the proposition.

included in the device, although the time management tools were well received. Ownership of the technology was shown to be important. Whilst the PDAs are loaned, students are reluctant to invest time and money in personalisation and extension. Universities and other institutions will need to provide students with more assistance in learning through personal technologies, including regular updates of timetables and content. It is difficult to commit much organisational resource for a small scale trial, but as more students bring their own devices into universities, change is now being driven by their demands as consumers.

The evaluation methods could not show what or how the students were learning. That is appropriate given the nature of the project, which was to explore the first use of a new technology (wireless PDA) over a long time period. It was not possible to predict in advance how students would use the devices, or even if they would adopt them at all. Since they interleaved use of PDAs with many other tools then it is not possible to factor out learning gains due to the PDAs. Indeed, the main purpose of providing them with the devices was to assist them in making their studies more organised and efficient, rather than to deliver core content. The results did show that some, but not all, students took the opportunity to organise their studies and to preview material using the PDAs. Most important, it did not suggest the need for a dedicated 'Mobile Learning Organiser', but rather for a device with communications facilities, a standard range of office and media tools, and access to learning content.

MyArtSpace

The MyArtSpace project also explored the adoption of novel technology, but in the more structured setting of a museum, to support curriculum learning (Vavoula, Meek, Sharples, Lonsdale, & Rudman, 2006; Sharples, Lonsdale, Meek, Rudman, & Vavoula, 2007; Vavoula, Sharples, Rudman, Lonsdale, & Meek, 2007). The aim of the project was to address a well-recognised problem (Guisasola, Morentin and Zuza, 2005) of lack of connection between a school museum visit and preparation and follow-up in the classroom.

In relation to Table 1, the learning was externally initiated, as part of the school curriculum, and it shifted from being externally managed, by a teacher in the classroom to learner managed, by the students in a museum or gallery. The duration was much shorter than the Mobile Learning Organiser project, comprising two classroom lessons and a museum trip. The settings were also more predictable, though the students were free to roam through the museum building. Although more constrained, the project posed a substantial challenge in that the evaluation had to inform the design of the MyArtSpace service and also to indicate benefits and problems for the learners, the schools and the museums.

MyArtSpace was a year-long project, funded by the Department of Culture Media and Sport to support structured inquiry learning between school classrooms and museums or galleries. Using the MyArtSpace service, children aged 11-15 can create their own interpretations of museum objects through descriptions, images and sounds, which they can share and present back in the classroom. Before the visit, the teacher in the classroom sets an open-ended question which the students should answer by gathering and selecting evidence from the museum visit. On arrival at the museum, students are given multimedia mobile phones which they can use to 'collect' exhibits (by typing a two-letter code shown next to museum exhibits which triggers a multimedia presentation on the phone), take photos, record sounds, or write text comments. This content is transmitted automatically by the phone to their personal online collection. Back at school, the students can view their collected

content on a web browser, organize it into personal galleries, share and present their findings in the classroom and then show the presentations to friends and family. Some 3000 children used the service at two museums and a gallery over the period of the project.

The evaluation team was fortunate in being involved throughout the project, from beginning to end. It was contracted to inform the design of the MyArtSpace service, which was being developed by a separate multimedia company, as well as to assess its educational value. To address this broad remit, we adopted a Lifecycle evaluation approach (Meek, 2006) that matches the evaluation method to the phase in the development lifecycle, providing outcomes that can feed forward to guide the next stages of development and deployment and also feed back to assist the design of new versions of the software.

The early stages of evaluation included stakeholder meetings with teachers, museum education staff and the software developers, to establish the goals and requirements of the service. These meetings proposed requirements (112 in total) that the stakeholders were asked to rate using the he MoSCoW technique from Dynamic Systems Development Method (Stapleton, 2003) to indicate that, for each requirement:

Must: must have this
Should: should have this if at all possible
Could: could have this if it does not affect anything else
Would: will not have this time, but would like to have in the future

Successive prototypes, starting with 'paper' designs, were given heuristic evaluations (Molich & Nielsen, 1990) whereby usability experts identified and rated usability problems. The prototypes were also assessed as to how they met each requirement. From the start, the evaluation covered the entire service, including teacher and museum support and training, so the teacher and museum guidelines, teacher information packs and training sessions were also assessed.

As the project moved from design and implementation to deployment, a series of studies were planned to assess the usability, effectiveness, and satisfaction of MyArtSpace. The team developed a three-level approach to evaluation.

Micro level: examined the individual activities that MyArtSpace enabled students to perform, such as making notes, recording audio, viewing the collection online, and producing presentations of the visit.
Meso level: examined the learning experience as a whole, exploring whether the classroom-museum-classroom continuity worked.
Macro level: examined the impact of MyArtSpace on educational practice for school museum visits.

For each level, the evaluation covered three stages, to explore the relationship between expectations and reality:

Stage 1: what was supposed to happen, based on pre-interviews with stakeholders and documentation including the teachers' pack.
Stage 2: what actually happened, based on observer logs, focus groups, and post-analysis of video diaries.
Stage 3: the gaps between findings from stages 1 and 2, based on reflective interviews with stakeholders and critical incident analysis of the findings from stages 1 and 2.

Taken together, the levels and stages provide a framework to evaluate usability (does the service do what was intended?), effectiveness (did the service support learning as expected, or were there unexpected benefits or problems?), and satisfaction (did the stakeholders find the service unexpectedly enjoyable or unpleasant?), with results that could be passed to systems designers, educators and policy makers. The specific evaluation methods included: one-to-one interviews with teachers; focus group interviews with students; video observations of a pre-visit lesson, museum visits and post visit lesson; attitude surveys; and telephone or email interviews with other stakeholders.

As a very brief summary of the results, at the micro level the system worked well, with the phones offering a familiar platform and the two letter code providing an easy way to activate multimedia in context. The transmission of data took place unobtrusively after each use of the photo, audio or note tool. The teachers indicated that their students engaged more with the exhibits than in previous visits and had the chance to do meaningful follow-up work.

At the meso level, a significant educational issue was that some students found difficulty in identifying, back in the classroom, pictures and sounds they had recorded. The time-ordered list of activities and objects they had collected provided some cues, but there is a difficult trade-off between structuring the material during the visit to make it easier to manage (for example by limiting the number of items that can be collected) and stifling creativity and engagement.

A significant issue emerged at the macro level. Although the system was a technical and educational success, there are significant barriers to wider deployment of a system like MyArtSpace. Many museums already provide audio guides and staff may be reluctant to spend time maintaining yet more technology. There is also the issue of who pays for the phone data charges: schools, museums, or students and their parents? The MyArtSpace service is now being marketed commercially as OOKL (www.ookl.org.uk) and has been adopted by some major UK museums and galleries that have the resources to support the service, with the company renting phone handsets to the venues with the software pre-installed. Wider adoption may depend on the next generation of mobile technology, when people carry converged phone/camera/media player devices that can easily capture everyday sights and sounds to a personal weblog (see also Pierroux in this volume). Then, the opportunity for schools will be to exploit these personal devices for learning between the classroom and settings outside school including field trips and museum visits.

Personal Inquiry

The Personal Inquiry (PI) project (http://www.pi-project.ac.uk/) has some similarities to MyArtSpace in that it connects learning in formal and informal settings, but there is a greater emphasis on providing a generic toolkit to support inquiry learning, starting in the classroom and then continued in a variety of settings including the school grounds, the city, homes, and science centres.

The project is a three-year collaboration between the University of Nottingham and the Open University, UK, to help young people aged 11-14 to understand themselves and their world through a scientific process of active inquiry across formal and informal settings. The children will use handheld and classroom computers to gather and assess evidence, conduct experiments and engage in informed debate. Their activities will be based around topic themes – Myself, My Environment, My Community – that engage young learners in investigating their health, diet and

fitness, their immediate environment and their wider surroundings. These topics are key elements of the new 21st century science curriculum (Millar & Osborne, 1998) that requires children to reason about the natural sciences as a complex system and to explore how people relate to the physical world.

The technology will be in the form of an inquiry learning toolkit running on small touch screen computer-phones, with integral cameras and keyboards, plus connected data probes, to enable learners to investigate personally-relevant questions outside the classroom, by gathering and communicating evidence. The toolkit will be designed to support scripted inquiry learning, where scripts are software structures, like dynamic lesson plans, that generate teacher and learner interfaces. These will orchestrate the learners through an inquiry learning process providing a sequence of activities, collaborators, software tools and hardware devices, while allowing the teacher to monitor and guide student activity. The children and their teachers will be able to monitor their learning activity, and to visualise, share, discuss and present the results, through a review tool accessible through a standard web browser running on a desktop or portable computer in the home or school. Teachers will also have a script authoring tool to create and modify the scripts, to support the learning of specific curriculum topics.

A challenge for evaluation is that the project needs to demonstrate the benefits, if any, of the general approach of scripted inquiry learning supported by mobile technology. The proposition is that the integrated system (mobile technology, inquiry methods, and learning between formal and informal settings) will provide the learning benefits, rather than any individual component. Thus, the learning benefits of each part cannot be tested separately, and the entire system is so different from traditional classroom teaching that there is little value to carrying out a comparative study of learning outcomes. Instead of assessing how children might learn *better* through scripted inquiry learning, the initial aim will be to assess how they learn *differently.* For the initial school trials we have adopted a critical incident study as one method of evaluation.

Over a two-week period of five science lessons, 30 students aged 14, planned a scientific investigation to explore the relation between heart rate and fitness (lesson 1) which they first explored in the relatively controlled environment of the classroom (lesson 2), then extended through a more active engagement with the inquiry process in the leisure centre (lesson 3), and concluded the work in the school library as they analysed the results (lesson 4) and created presentations (lesson 5). All the teaching sessions were videotaped with three cameras: one fixed camera giving an overview of the lesson and two others to record closer views on the classroom or group activity. Radio microphones were used to provide good sound quality.

A critical incident analysis of the videotapes identified specific learning breakthroughs and breakdowns (Sharples, 1993). Breakthroughs are observable critical incidents which appear to be initiating productive new forms of learning or important conceptual change. Breakdowns are observable critical incidents where a learner is struggling with the technology, is asking for help, or appears to be labouring under a clear misunderstanding. They may either be predictable (e.g. the intervention may be aimed at producing conceptual change) or unpredicted (e.g. a child uses the technology in novel ways, or makes an unforeseen connection or conceptual leap). The critical incident analysis was conducted as follows: the videotapes were separately viewed by three researchers to identify obvious and informative breakdowns or breakthroughs (for example, where there is some activity and discussion on the video to indicate causes or solutions to the problem, or that suggest the nature of the learning); the identified critical incidents were then

compared to reach an agreed set of incidents that might inform design; and a videotape of the selected incidents was also played to a focus group of students, to elicit their interpretation of the events.

As a result of this process we identified eight incidents (four breakdowns, three breakthroughs and one incident that could be interpreted as both a breakthrough and a breakdown). An example of a breakthrough was the teacher herself wearing a monitor, continually generating a graph of her heart-rate on the class display which she referred to during the lesson. An example of both a breakthrough and a breakdown came in the fitness centre where the children were able to view and discuss their data as it emerged, successfully creating a *micro-site* for learning, but the software did not indicate on the graph where a fitness exercise started and ended.

The PI project is still continuing, with further trials planned to connect learning in the classroom, homes and city centres. These will present new problems in evaluation, in particular the practical and ethical problems of conducting evaluations in a home. We are developing ethical guidelines to cover this type of mobile learning evaluation, including: ensuring that the children are fully informed about how their learning activities outside the classroom may be monitored, allowing children to decide where and when to collect data, and ensuring that material captured and created by the children will be subject to normal standards of copyright and fair use, so that inappropriate material will be deleted and the authors of the teaching materials and field data retain copyright and moral rights of authorship over their material.

Summary

Evaluation of mobile learning poses particular challenges not only because it introduces novel technologies and modes of learning, but also because it can spread across many contexts and over long periods of time. It is generally not possible to control factors to an extent that would make a comparative study appropriate. However, it may be worth attempting such a study when there is a well-defined learning activity and a comparative less-expensive technology, for example on a field trip to compare learning supported by PDAs with a similar trip using paper worksheets and children's own phone cameras.

To meet the challenges, researchers are developing distinctive methods of evaluation that are sensitive to time and context. A first step in planning a mobile learning evaluation is to determine whether it is concerned with technology development, or appropriation, or the implications of new or existing mobile technology for learning. Technology development can be guided and evaluated by a variety of human-computer interaction methods, though attention will need to be paid to how the technology performs in realistic settings, such as outdoor sunshine. Technology appropriation was discussed in the Mobile Learning Organiser section, with methods that include diary-interview and periodic surveys. These can reveal changing patterns of use and interest, but not the processes and outcomes of learning. Waycott (2004) presents a valuable framework for analysing the appropriation of mobile technologies for learning, derived from case studies.

In evaluating learning with mobile technology it may be useful to start with Table 1 to determine whether the learning is initiated and managed by the learner, or others. Mobile learning that is self-initiated and managed (for example, long-term language learning, or learning on vacation) is unlikely to be predicable either in content or context. Capturing evidence of the learning may be difficult, particularly if it spans

multiple technologies. Vavoula has developed a successful method to study informal mobile learning based on structured diaries kept by learners, followed up by interviews (Vavoula, 2005). This is a labour-intensive process but it has revealed contexts and conditions of mobile learning. Another possibility is to phone or text the learner at pre-agreed intervals to ask about current or recent learning activities.

For learning that is either externally initiated or managed, the contexts and topics are more likely to be pre-determined, so there are likely to be opportunities to examine teaching materials and settings in advance and plan where and how to observe the learning. Data capture methods include videotaped observations of individuals or groups (preferably wearing radio microphones for good sound quality), log files of human-computer interactions, and observer notes. Analysis of the data can include critical incident methods (including interviews with participants to discuss replays of the incidents), interaction or discourse analyses, and analysis of log data (possibly synchronised to videotapes) to reveal changing patterns of interaction, for example as the learner alternates between engagement in a learning activity and reflection on findings.

Lastly, learning that is both externally initiated and structured (for example, use of handheld technologies in a classroom) can be evaluated through a variety of methods, including those above, as well as learning outcome measures and comparative studies.

Evaluation of mobile learning is not intrinsically different to other forms of learning, in that we want to understand the individual and collective processes of coming to know and the resulting changes in knowledge, skill and experience. This chapter has suggested some ways in which those processes can be observed and analysed across contexts and long time periods.

Acknowledgements
The Mobile Learning Organiser was a project of the Educational Technology Research Group, Electronic, Electrical and Computer Engineering, University of Birmingham. Susan Bull, Tony Chan and Dan Corlett contributed to the project and its results.

Peter Lonsdale, Julia Meek, Paul Rudman and Giasemi Vavoula contributed to the evaluation of MyArtSpace. The project was funded by the Department for Culture, Media through Culture Online. The Myartspace was designed and developed by The Sea (http://www.the-sea.com) and is now marketed as OOKL (www.ookl.org.uk). We thank the students, teachers, and museum educators who took part in our trials.

The Personal Inquiry project acknowledges support from the Economic and Social Research Council and the Engineering and Physical Sciences Research Council UK through the Teaching and Learning Research Programme. The design and evaluation discussed here were carried out by Stamatina Anstopoulou, Steve Benford, Charles Crook, Chris Greenhalgh, Claire O'Malley, Hazel Martin and Michael Wright. We are grateful for the support and engagement of the teacher and children who participated in this study. The project is a collaboration between the University of Nottingham and the Open University, UK.

References

Benedek, J., & Miner, T. (2002). *Measuring desirability: New methods for evaluating desirability in a usability lab setting.* Paper presented at the UPA 2002 Conference, July 8-12, 2002., Orlando, FL.

Corlett, D., Sharples, M., Bull, S., & Chan, T. (2005). Evaluation of a mobile learning organiser for university students. *Journal of Computer Assisted Learning, 21*(3), 162-170.

Damala, A., & Lecoq, C. (2005). Mobivisit: Nomadic Computing in indoor cultural settings. A field study in the museum of Fine Arts, Lyon. In X. Perrot (Ed.), *ICHIM International Cultural Heritage Informatics Meeting September 21-23, 2005, Paris, France*.

Guisasola, J., Morentin, M., & Zuza, K. (2005). School visits to science museums and learning sciences: a complex relationship. *Physics Education, 40*(6), 544-549.

Livingstone, D. W. (2001). *Adults' Informal Learning: Definitions, Findings, Gaps and Future Research* (No. Working paper 21). Toronto: NALL (New Approaches to Lifelong Learning).

Meek, J. (2006). *Adopting a Lifecycle Approach to the Evaluation of Computers and Information Technology.* Unpublished PhD Thesis, The University of Birmingham, UK.

Millar, R., & Osborne, J. (1998). *Beyond 2000: Science education for the future*: King's College, University of London.

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33*(3), 338-348.

Naismith, L., Sharples, M., & Ting, J. (2005). *Evaluation of CAERUS: a Context Aware Mobile Guide*. Retrieved 8th May 2008, from http://www.mlearn.org.za/CD/papers/Naismith.pdf

O'Malley, C., Vavoula, G., Glew, J.P., Taylor, J., Sharples, M., Lefrere, P., Lonsdale, P., Naismith, L. & Waycott, J. (2005). Guidelines for learning/teaching/tutoring in a mobile environment. MOBILearn project report D4.1. Retrieved 13th August 2008, from http://www.mobilearn.org/download/results/public_deliverables/MOBIlearn_D4.1_Final.pdf.

Rogers, Y., Price, S., Fitzpatrick, G., Fleck, R., Harris, E., Smith, H., et al. (2004, June 1-3). *Ambient wood: designing new forms of digital augmentation for learning outdoors.* Paper presented at the 2004 conference on Interaction design and children: building a community (IDC 2004), Maryland, USA.

Schwabe, G., & Goth, C. (2005). Mobile learning with a mobile game: design and motivational effects. *Journal of Computer Assisted Learning, 21*(3), 204-216.

Sharples, M. (1993). A Study of Breakdowns and Repairs in a Computer Mediated Communication System. *Interacting with Computers, 5*(1), 61-77.

Sharples, M., Taylor, J., & Vavoula, G. (2007). A Theory of Learning for the Mobile Age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage Handbook of Elearning Research* (pp. 221-247). London: Sage.

Sharples, M., Lonsdale P., Meek J., Rudman P.D., Vavoula G.N. (2007). An Evaluation of MyArtSpace: a Mobile Learning Service for School Museum Trips. In A. Norman & J. Pearce (eds.) *Proceedings of 6th Annual Conference on Mobile Learning, mLearn 2007*, Melbourne. Melbourne: Universitty of Melbourne, pp. 238-244.

Stapleton, J. (2003). *DSDM: A Framework for Business-Centered Development*. Boston, MA: Addison-Wesley Longman Publishing Co.

Traxler, J., & Kukulsa-Hulme, A. (2005). *Evaluating mobile learning: Reflections on current practice.* Paper presented at the mLearn 2005, Cape Town.

Valdivia, R., & Nussbaum, M. (2007). Face-to-Face Collaborative Learning in Computer Science Classes. *International Journal of Engineering Education, 23*(3), 434-440.

Vavoula, G., Meek, J., Sharples, M., Lonsdale, P., & Rudman, P. (2006). A lifecycle approach to evaluating MyArtSpace. In S. Hsi, Kinshuk, T. Chan & D. Sampson (Eds.), *Proceedings of the 4th International Workshop of Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE 2006), Nov 16-17, Athens, Greece.* (pp. 18-22): IEEE Computer Society.

Vavoula, G. N. (2005). *D4.4: A Study of Mobile Learning Practices, Report of MOBILearn project*. Retrieved 8th May, 2008, from www.mobilearn.org/download/results/public_deliverables/MOBIlearn_D4.4_Final.pdf

Vavoula, G., Sharples, M., Rudman, P., Lonsdale, P., Meek, J. (2007). Learning Bridges: a role for mobile technologies in education. *Educational Technology*,Vol. XLVII, No. 3, May-June 2007, pp. 33-37.

Waycott, J. (2004). *The appropriation of PDAs as learning and workplace tools: an activity theory perspective.* Unpublished PhD Thesis, The Open University, Milton Keynes.