Meeting the Challenges in Evaluating Mobile Learning: A 3-level Evaluation Framework

Giasemi Vavoula
Department of Museum Studies
University of Leicester, UK
gv18@le.ac.uk

Mike Sharples
Learning Sciences Research Institute
University of Nottingham, UK
mike.sharples@nottingham.ac.uk

ABSTRACT

We propose six challenges in evaluating mobile learning: capturing and analysing learning in context and across contexts, measuring mobile learning processes and outcomes, respecting learner/participant privacy, assessing mobile technology utility and usability, considering the wider organisational and socio-cultural context of learning, and assessing in/formality. A three-level framework for evaluating mobile learning is proposed, comprising a micro level concerned with usability, a meso level concerned with the learning experience, and a macro level concerned with integration within existing educational and organisational contexts. The paper concludes with a discussion of how the framework meets the evaluation challenges and with suggestions for further extensions.

Keywords

Mobile learning evaluation; learning context; evaluation framework; learning outcomes; ethics; informality and formality of learning; requirements for evaluation.

Introduction

Mobile learning is a relatively new research area, with the first research projects appearing in the second half of the 1990s and the first international research conferences less than a decade ago. It is a field whose practice has not yet been standardised in terms of research frameworks, methods and tools. Thankfully, mobile learning has a lot of common ground with related research areas including Technology-Enhanced Learning (TEL) and Mobile Human-Computer Interaction (mobileHCI). 'Borrowing' frameworks and methods from these areas has been common practice for early mobile learning research, providing researchers with useful starting points.

As our conceptions and understanding of mobile learning deepen, these 'borrowed' frameworks and tools might no longer be adequate. We now appreciate mobile learning not just as learning that is facilitated by mobile technology, but also as the processes of coming to know through conversations and explorations across multiple contexts amongst people and personal interactive technologies (Sharples *et al.* 2007a). Such evolving conceptions introduce challenges to all aspects of mobile learning research, including evaluation. As the field matures, our frameworks and tools need to address these challenges.

In this paper we summarise six challenges in evaluating mobile learning: capturing and analysing learning in context and across contexts, measuring the processes and outcomes of mobile learning, respecting learner/participant privacy, assessing mobile technology utility and usability, considering the wider organisational and socio-cultural context of learning, and assessing in/formality. The paper proposes an evaluation framework with three levels: a micro level concerned with usability, a meso level concerned with the learning experience, and a macro level concerned with integration within existing educational and organisational contexts. The paper demonstrates how this framework has guided data collection and analysis in one mobile learning evaluation project, and concludes with a discussion of how it meets the evaluation challenges and with suggestions for further extensions.

Challenge 1: Capturing learning context and learning across contexts

A major task for educational evaluation is to identify and analyse learning within and across contexts. For mobile learning, the interest is not only in how learning occurs in a variety of settings, but also how people create new contexts for learning through their interactions and how they progress learning across contexts. This poses a significant challenge to evaluators of mobile learning. In order to establish, document and evaluate learning within and across contexts, a researcher needs to analyse: the physical setting and the layout of the learning space (where); the social setting (who, with whom, from whom); the learning objectives and outcomes (why and what); the learning methods and activities (how); the learning progress and history (when); and the learning tools (how).

When evaluating learning in a traditional classroom, researchers generally have access to information about these context elements before, during and after the learning experience. Thus, they can inspect the classroom and interview the teacher and learners in advance of a lesson to discover the objectives, methods, lesson plan and tools. To evaluate a school museum visit or field trip, the researcher can visit the site and inspect the lesson plan, but will generally not know in advance the route that each student will take. For personal or family visits to museums or other learning sites, neither the objectives nor the trajectory may be known in advance. Learning objectives may arise as a response to interactions with the environment and learning trails may be guided by curiosity or unplanned events. The learners themselves may not be known in advance, for example when evaluating the learning experience of museum visitors randomly selected at the museum entrance. Personal mobile learning embraces any learning event where people, individually and collectively, continually create micro-sites for learning out of the available physical and social resources. In considering this generic case, the setting, objectives, methods and processes may all be unpredictable.

Table 1 below portrays the increasing vagueness in moving from evaluating a classroom lesson, to a school museum visit, to personal or family museum visits, to personal mobile learning across formal and informal settings. Each set of context elements requires specific evaluation methods, to match the actual learning processes and outcomes to expectations, or to capture contingent and unexpected learning events.

-- _____vagueness_____ ++

	Classroom	**School museum visit or field trip**	**Personal or family visit**	**Personal mobile learning**
Physical setting	☑ Conventional and static	☒ Moving around a fixed location	☒ Moving around a fixed location	☒ Unpredictable & changing
Social setting	☑ Fixed	☑ Pre-arranged	☑ Pre-arranged	☒ Unpredictable and changing
Learning objectives and outcomes	☑ Externally set	☑ Externally set	☒ Personally set or contingent	☒ Personally set or contingent
Learning method and activities	☑ Pre-determined	☑ Pre-determined	☒ Pre-determined or contingent	☒ Pre-determined or contingent
Learning progress and history	☑ Pre-determined	☒ Pre-determined or contingent	☒ Mostly contingent	☒ Contingent
Learning tools	☑ Provided	☑ Provided by school or museum	☒ Provided & personally owned	☒ Personal & serendipitous

Table 1. Context elements relevant to the learning researcher.

Recent research efforts have focused on devising tools and methods appropriate for capturing and analysing mobile learning contexts. Some efforts concentrate on implementing technology-based solutions for data collection, such as mobile eye tracking or wearable interaction capture kits (Roto *et al.* 2004). Although these have the advantage of capturing accurate data in context, they have some disadvantages, not least the obtrusiveness of the apparatus used. Other efforts opt for cooperative inquiry-based solutions (Hsi 2007), such as using learners' accounts of the experience through retrospective interviews, diaries, or attitude surveys (Clough & Jones 2006; Vavoula 2005). These have different shortcomings such as the accuracy of recall, the degree to which post-rationalisation skews data, and the effect of the participants' concern over the image they project.

Increasingly, mobile evaluation designs include mixed methods. These are useful not only for validating data, but also for capturing different perspectives of the learning experience. Thus, collected data might include recorded video, audio transcripts, observation notes, artefacts produced by the learners, and application screenshots. Interpreting such rich collections of data can be challenging too, in terms of assembling it into a meaningful, accurate and elaborate account of the learning experience. Related research addresses the design of tools and methods to support the sequencing, synchronisation, inter-relation and visualisation of evaluation data (Greenhalgh *et al.* 2007; Papadimitriou *et al.* 2007; Smith *et al.* 2007).

Challenge 2: Has anyone learned anything?

A second challenge that faces mobile learning evaluation is the assessment of learning processes and outcomes. In traditional learning settings such as the classroom there are well-established and accepted methods for the assessment of learning activities, such as essay writing, multiple choice tests, open-book exams, and unseen examinations. Distinctions have been made between formative assessment (aiming to provide students with feedback regarding their progress) and summative assessment (aiming to judge and sum up the students' achievements) (Scriven 1967), with formative assessment bearing the greater potential to aid and complement teaching and learning (Black & Wiliam 1998a; 1998b).

Summative assessment is often used as a measure of success of the teaching as well as a measure of effectiveness of the learning (Boud 1995), but with many (often unresolved) issues regarding the reliability and validity of summative assessment methods (see Knight 2001 for a discussion of these issues). Despite these difficulties, summative assessment can be meaningful in formal learning contexts where learning objectives and desired outcomes are well specified in advance. By contrast with formal education, mobile, informal learning can be both personal and elusive. The learning may be personally initiated and structured, such that it is not possible to determine in advance where the learning may occur, nor how it progresses or what outcomes it produces. It may also be difficult to track the progress of learning if it occurs across multiple settings and technologies.

An alternative approach is to examine the experience for evidence which might suggest that productive learning is taking place. For example, in the context of museum learning, Griffin and Symington (1998) suggest to watch for instances where learners initiate and show responsibility for their own learning (e.g. by writing, drawing, or taking photos by choice; deciding where and when to move), are actively involved in learning (e.g. by absorbed, close examination of resources; or persevering with a task), make links and transfer ideas and skills (e.g. by comparing evidence), and share learning with experts and peers (e.g. by talking and gesturing; or asking each other questions). Adaptations of the Critical Incidents method (Flanagan 1954) provide one way to achieve this. For example, activities of learners who wear radio microphones are videotaped at a discrete distance. The evaluators then watch the videotapes to identify observable critical incidents that appear to be breakthroughs (indicating productive new forms of learning or important conceptual change) or breakdowns (where a learner is struggling with the technology, is asking for help, or appears to be labouring under a clear misunderstanding). These incidents can be assembled into a compilation tape and reviewed with the learners for further elaboration (Vavoula *et al.* forthcoming) or analysed as is (Anastopoulou *et al.* 2008).

Another alternative is to focus on learner perceptions of the learning experience rather than learning outcomes in terms of cognitive gains. Attitude surveys have been used extensively in the mobile learning literature to measure learner attitudes towards the technology and their enjoyment of the experience. Since attitudes are closely related to intrinsic motivation and learning agency (Hidi & Harackiewicz 2000), they can be a reliable predictor of conditions for effective learning (though not necessarily of learning outcomes). However, the mobile learning community has yet to produce standardised attitude measurement instruments such as those available in other fields (e.g. science learning - Moore & Sutman 1970).

Information useful in assessing learning can also be found in learner-created artefacts, such as log files of computer activity or web access, the results of online quizzes, learner-created media, and personal reflective documents such as blogs and e-portfolios. Further work is needed to integrate these into a revealing and valid assessment of learning.

The challenge of assessing learning is not unique to mobile learning and is not easily solved. Although a learning experience can be a well defined event with a start and a finish, learning is an ongoing, lifelong process of personal transformation and, as such, requires longitudinal, historical assessment.

Challenge 3: An ethical question

Research ethics frameworks have governed research involving human subjects for decades. With the increasing use of the Internet as the research object and medium, accounts of virtual research ethics prevailed, along with analysis of differences in the nature of ethical issues confronted by virtual versus traditional research (Buchanan 2004). The challenge for mobile learning evaluation is to translate these issues into ethical guidelines appropriate for mobile contexts.

The evaluation of mobile learning presents particular ethical problems, beyond those routinely associated with a study of people and technology, i.e. ensuring their safety and informed cooperation. A fundamental need is to explicate the purpose and principles of the evaluation within an ethical framework. Ethics can arise from differing philosophical foundations regarding the nature of 'reality' and the value of the scientific method in validating claims by analysis of objective data.

> Within a modernist framework the researcher strives for objectivity. Engaging in postmodern research involves researchers reflexively asking why they are doing this research in this way, what it is silent about, what gives it authority, and who is privileged by it (Traxler & Bridges 2004:204).

As studies of learning move out of the classroom into homes and outdoor locations, so the evaluation will need to rely more on a combination of data collected automatically by mobile devices (such as logs of user interaction, time and location) and self-reports from learners. These do not fit naturally together, since they exemplify objectivist and postmodern approaches to the study of learning. For example, a current study (as yet unpublished) of children using mobile phones to create records of their daily eating habits has found that some children deliberately avoid photographing unhealthy food items. This is not simply a matter of treatment of missing data items, but indicates deeper problems of children's self-image, research by proxy using mobile devices, intrusion into daily life, power and willing cooperation, and the interpretation of silence. As mobile learning grows in scale and scope, evaluators must address the reliability of evidence (particularly when collected outside the lab or classroom) and difficulties of conducting scientifically rigorous studies as a basis for formulating evidence-based policy[1].

[1] We are grateful to a reviewer for bringing this aspect of evaluation to our attention.

Traxler and Bridges indicate specific ethical issues of evaluating mobile learning, including: explaining the scope of mobile learning in a succinct and appropriate way, gaining informed consent for novel forms of interaction (such as learning by SMS), and identifying the contribution of learners across multiple devices and contexts. Other issues include identifying the ownership of material collected across contexts (such as field trips), the rights of participants to know when and how they are being monitored during their daily lives, and possible health dangers associated with regular use of wireless technologies (Ahonen 2008; Patrick *et al.* 2008).

Obtaining informed consent can be problematic: the previous sections described the vagueness of mobile learning context and the elusiveness of mobile learning outcomes. When evaluators are uncertain of what will constitute the mobile learning experience, how accurately can they inform the participants of what data is sought and why? Assuming that a vague description of the requirements for participation is acceptable, how can learners consent to disclosing information about events they currently do not know when, where and under what circumstances will take place?

Even if the essence of the evaluation is successfully conveyed to the participants, and they consent to it, there are still important issues to consider relating to the degree to which they will co-operate in practice – either in terms of disclosing all that might be relevant, or in terms of carrying out related practical tasks such as synchronising their mobile devices as and when requested (Trinder *et al.* 2007).

Furthermore, when asking participants to record their own learning (either through wearing/carrying recording equipment or by keeping a written record of their learning), we are in essence making them co-researchers. In doing so, we also need to ensure that they are able themselves to follow the ethics rules and regulations; and are more legitimate research partners than inactive participants (Bakardjieva *et al.* 2004).

A major challenge then for mobile learning evaluation is to accurately inform participants, to ease their participation, and to build capacity in ethical research and mobile learning practice by providing them with appropriate ethics training. In the process evaluators should ask themselves how much they really need to know, and investigate best practices in safeguarding and disseminating sensitive personal data.

Challenge 4: Mobile technology

Evaluations of mobile learning often reference inherent limitations of mobile devices, such as their small screens, short battery lives, intermittent connectivity, and associated human factors, all of which affect their usability (see e.g. Corlett *et al.* 2005). As the focus of research shifts from the mobility of the technology to the mobility of the learner, additional issues arise as learners move across multiple, quickly-dating devices, both personal and public, possibly over short time periods in multiple locations. Extracting learning interactions from this mesh of technology interactions requires synchronization of data capture and analysis across multiple devices and interfaces. Assessing the usability of the mobile technology and the effectiveness of its integration with the mobile learning practice remains a high priority for evaluation.

Thus, challenges of mobile human-computer interaction stemming from the complexity introduced by physical movement and changing variables (Kjeldskov & Stage 2004) and the small scale and ubiquitous nature of mobile devices (Hagen *et al.* 2005), add to the challenges already facing mobile learning evaluation.

Challenge 5: Seeing the bigger picture

There is a wealth of literature on the relation between information technology and institutional change (e.g. Fountain 2001) and methods of performance management such as the balanced scorecard (Kaplan & Norton 1996) and Six Sigma (Pande *et al.* 2000). Becta, the UK Government agency leading the introduction of technology in education, provides a strategy for the introduction and effective use of technology across all education sectors. This describes the progression of an institution in confidence with technology from 'enabled', through 'capable' to 'confident' through introduction of appropriate infrastructure, planning and leadership, learner access to resources, and personalisation of learning. Its Performance Framework (BECTA 2008:47) indicates the systemic changes needed to achieve the goals of improved personalised learning experiences. These are: confident system leadership and innovation, technology confident effective providers, engaged and empowered learners and enabling infrastructure and processes.

For Higher Education (HE), Oliver and Harvey (2002) suggest four kinds of impact of educational technologies: impact on students' learning, impact on individual academics' practice, impact on institution, and national impact. Also in the context of HE, Price and Oliver (2007) identify three types of impact studies: anticipatory, ongoing and achieved. Anticipatory studies relate to pre-intervention intentions, opinions and attitudes; ongoing studies focus on analysing processes of integration; and achieved studies are summative studies of technology no longer 'novel'. Riley (2007) extends this impact framework by distinguishing between minor modifications and culturally significant changes in practice, and suggesting that different kinds of change will emerge over different timescales.

Mobile learning evaluation has similar issues regarding impact. It needs to examine how learning takes place within a personal, socio-cultural and institutional context, to chart the progression of institutions in their maturity of support for learning with mobile technology, and examine the relation between personal and institutional learning. It needs to address how the immediate learner experience within these contexts blends with or confronts existing practices to lead to new practices, by analysing this change process over extended periods of time. These requirements necessitate an extended view of the role of evaluation as a continual process of adjustment and fine-tuning.

Challenge 6: Formal or informal?

Mobile learning is often defined in terms of the technology that mediates the learning experience: if the technology is mobile, so is the learning. Mobility, however, is not an exclusive property of the technology, it also resides in the lifestyle of the learner, who in the course of everyday life moves from one context to another, switching locations, social groups, technologies and topics; and learning often takes place inconspicuously or is crammed in the short gaps between these transitions. Although this view of learning is inclusive of formal education contexts, it is particularly pertinent to everyday, informal learning.

Nevertheless, characterising a learning experience as formal or informal can be complicated. For example, is the learning of pupils visiting a museum (largely considered an informal learning setting) with their school (an irrefutably formal learning setting) a case of formal or informal learning? There is a large literature related to definitions of informal learning and related terminology, a review of which is beyond the scope of this paper. However, a general tendency seems to be to define informal learning in contrast to formal learning, and formal learning in turn to be confined to learning that takes place in educational settings. Colley et al. (2003) argue that "seeing informal and formal learning as fundamentally separate results in stereotyping and a tendency for the advocates of one to see only the weaknesses of the other … It is more sensible to see attributes of informality and formality as present in all learning situations". They advocate four groups of attributes: those related to the learning process, to the location and setting, to the learning purposes, and to the learning content. They propose that attributes of in/formality are interrelated in different ways in different

learning situations, and that those attributes and their interrelationships influence the nature and effectiveness of learning in any situation.

Understanding such attributes of in/formality and their interrelationships in mobile learning is important for evaluation. It is not only a case of analysing pre-existing practices in terms of processes, settings, purposes and content, but also of capturing how the introduction of mobile learning practices, or new ways of supporting them, can change the in/formality of the learning experience.

Precepts for mobile learning evaluation

The challenges discussed in the previous sections translate into a set of basic precepts for mobile learning evaluation:

P1. Capture and analyse learning in context, with consideration of learner privacy (challenges 1, 3)

P2. Assess the usability of the technology and how it affects the learning experience (challenge 4)

P3. Look beyond measurable cognitive gains into changes in the learning process and practice (challenge 2)

P4. Consider organisational issues in the adoption of mobile learning practice and its integration with existing practices and understand how this integration affects attributes of in/formality (challenges 5, 6)

P5. Span the lifecycle of the mobile learning innovation that is evaluated, from conception to full deployment and beyond (challenges 1-6)

As an illustration of how these might guide evaluation in practice, the following section presents an evaluation framework for mobile learning and its application in the context of the Myartspace project.

M3: A three-level framework for evaluating mobile learning

Myartspace supports structured inquiry learning through technology that connects learning in the classroom with learning in museums and galleries. Detailed descriptions of the project and the evaluation process and outcomes have been presented elsewhere (Sharples *et al.* 2007b; Vavoula *et al.* 2006a; Vavoula *et al.* 2007; Vavoula *et al.* 2006b). In summary, Myartspace addresses the problem of how to connect a school museum trip with classroom activities of planning and further study. It enables school students to create their own interpretations of museum exhibits through descriptions, images and sounds they collect at the museum, which they then review, reflect upon and share outside the museum. Before the visit, the teacher will typically set an open-ended question that the students can answer by gathering and selecting evidence from the museum visit. On arrival at the museum, students are given multimedia mobile phones running custom software which they can use to collect multimedia presentations of exhibits, take photos, record sounds, or write text comments. This content is transmitted by the phone into a time-ordered collection on their personal web space. Back at school, the students can organise the material into online galleries to present their findings in the classroom and share with their friends and family.

M3, the evaluation framework developed in the context of Myartspace, followed the Lifecycle approach to educational technology evaluation proposed by Meek (2006). This places evaluation at the centre of the development process, from the early stages of design to a final assessment of the deployed technology in use. The Lifecycle approach can be matched to a sequential systems development process (Royce 1970), with evaluations undertaken at the end of each stage, or to iterative (Larman & Basili 2003) or socio-cognitive methods (Sharples *et al.* 2002) where evaluation activities are undertaken at key points that are of most value to support the design process or inform stakeholders, with the outcomes of each evaluation guiding the next phase of the system development or feeding into an iteration of an earlier phase.

Evaluation under M3 is conducted at three levels:

1. *Micro level*, which examines the individual activities of the technology users and assesses the usability and utility of the educational technology system. For Myartspace the activities included collecting objects through exhibit codes, making notes, contacting people who had collected a particular item, recording audio, and taking pictures.
2. *Meso level*, which examines the learning experience as a whole, to identify learning breakthroughs and breakdowns. It also assesses how well the learning experience integrates with other related activities and experiences. For Myartspace, evaluation at this level involved exploring whether there was a successful connection between learning in the museum and the classroom, as well as identifying critical incidents in the museum that reveal new patterns and forms of learning or where learning activity is impeded.
3. *Macro level*, which examines the impact of the new technology on established educational and learning practices and institutions. For Myartspace this related to the organisation of school museum visits. The evaluation at this level examined the appropriation of the new technology by teachers, the emergence of new museum practices in supporting school visits, and how these related to the original project visions.

The development of Myartspace comprised four broad phases: (1) Requirements analysis, to establish the requirements for the socio-technical system (the users and their interactions with technology) and specify how it would work, through consultation with the different stakeholder groups; (2) Design of the user experience and interface; (3) Implementation of the service; and (4) Deployment of the service. These are compatible with an Agile Development approach (Beck *et al.* 2001) where requirements can evolve throughout the development process to take account of the evaluations of usability, learning effectiveness and institutional adoption. Thus, the requirements analysis persisted throughout the project lifecycle, and covered all three levels of analysis (micro, meso and macro).

Figure 1 illustrates this gradual introduction of evaluation activities at the three framework levels over all project phases. The horizontal axis in Figure 1 depicts time; the change of focus development phase over time is shown, as is the persistence of requirements analysis throughout the project lifecycle. The shaded areas represent activities for requirements analysis (dark gray) and evaluation at the three levels during design, implementation and deployment (all other shades).

The emphasis on level of requirements analysis changes during the development process. At the start of a project, the requirements analysis must take account of all levels to set initial requirements for an educational experience that integrates technology, effective learning and institutional support. As the project progresses, the technology matures, so that changes to requirements become focused on the learning context and institutional adoption. At the end of the project, the requirements have been finalised and are evaluated at all levels.

The emphasis on level of evaluation also changes during the development process. Early evaluations at micro level inform the user interface and human-technology interactions. Once the technology is robust enough to allow assessment of educational value, evaluation activities at the meso level are introduced during the implementation phase. Similarly, the macro level requires that the technology is in place and used for long enough to establish its effects on e.g. school museum visiting practice, so evaluation activities at the macro level may be introduced during the deployment phase.

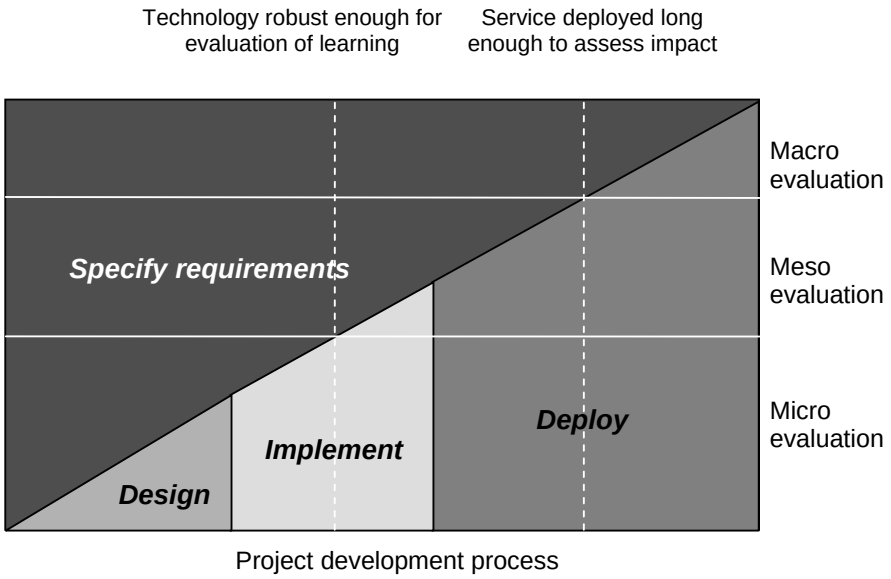

Figure 1: Evaluation activities at the three levels over the project phases.

To establish the value of the service at each of the three levels, evaluation activities explore the gap between expectations and reality and uncover unforeseen processes and outcomes. This is enacted in two stages of data collection and a stage of analysis:

1. Stage 1: collect data about what is supposed to happen at a level. User expectations at each level can be captured through interviews with users (e.g. teachers, students, museum staff) and by analysing technical requirements specifications, user documentation, training sessions and lesson materials.
2. Stage 2: collect data about what actually happened at a level. The user experience is documented through observations and video and audio recordings to establish the reality of technology use for the different users.
3. Stage 3: examines the gaps between user expectations and reality through a combination of reflective interviews with users and critical analysis of the data collected in stages 1 and 2.

In summary, M3 follows a Lifecycle approach of continuous strategic evaluation to guide an agile approach to software development and to inform stakeholders in the development process. It assesses the evolving design and implementation at three levels, of usability, educational effectiveness and institutional adoption. For each level the evaluation relates what should happen (through interviews with stakeholders and examination of documents) to what actually happens (through observation of user experience) and examines any gaps between expectation and reality as evidence of a need to modify requirements, design, implementation, or deployment. These findings guide the next phase of the system development or feed into an iteration of an earlier phase.

Table **2** summarises requirements analysis and evaluation activities and the respective data collection methods in the Myartspace project at each level of M3, for all project phases.

Data collection for requirements and evaluation activities	**Level**	**Phase**
Requirements analysis		
Stage 1: 'expectations' data collection		
• Scoping study of previous projects and related recommendations • Consultation workshop on 'User Experience' to establish requirements	All	Requirements
Stage 2: 'reality' data collection		
• Data supplied by evaluation analysis	All	Requirements
Heuristic Evaluations (examining how system designs compare to expectations re established design heuristics)		
Stage 1: 'expectations' data collection		
• Established design heuristics	Micro	All
Stage 2: 'reality' data collection		
• Experts undertaking heuristic evaluation	Micro	All
Technical testing prior to trials		
Stage 1: 'expectations' data collection		
• Data supplied by system requirements	Micro	Implement
Stage 2: 'reality' data collection		
• System performance tests	Micro	Implement
Full scale user trial (Key Stage 2 class visits Myartspace museum)	All	Implement/Deploy
Stage 1: 'expectations' data collection		
• Examine system documentation (Teacher's Pack and Lesson Plans, online help) for descriptions of functionality • Interview teacher prior to lesson to assess level of knowledge and expectations for functionality • Observe training sessions at museum and school to document how functionality is described to teachers/students. • Student questionnaires regarding expectations of system functionality in forthcoming lesson	Micro	Implement/Deploy
• Analyse description of educational experience based on Teacher's Pack and Lesson Plans • Interview teachers and museum educators prior to lessons about what they have planned for the students' learning experience • Observe teachers and museum educators while presenting learning experience to students in the classroom/museum • Student questionnaires regarding expectations of learning experience in forthcoming lesson	Meso	Implement/Deploy
• Analyse descriptions in service promotion materials, original proposal, minutes of early project meetings • Interviews with stakeholders to elicit initial expectations for impact of service	Macro	Deploy
Stage 2: 'reality' data collection		
• Observe lesson to establish actual teacher and student experience of functionality • Interview teacher after the lesson to clarify experience of functionality • Questionnaire and focus groups with students after the lesson to capture experience of functionality	Micro	Implement/Deploy
• Observe educational experience in museum/classroom o Note critical incidents that show new forms of learning or educational interaction o Note breakdowns • Interviews/focus groups with teachers, museum educators, students on educational experience in museum/classroom	Meso	Implement/Deploy
• Review of press coverage and interviews with stakeholders to document impact/transformations effected by the service	Macro	Deploy

Table 2: Methods used for data collection during requirements analysis and evaluation activities at each level, for each project phase

Data analysis for requirements and evaluation activities	Level	Phase
Requirements analysis		
Stage 3: data analysis		
• Workshop to finalise educational and user requirements • Revisions of requirements in light of evaluation findings	All	Requirements
Heuristic Evaluations		
Stage 3: data analysis		
• Analysis of expert reports and production of (re)design recommendations	Micro	All
Technical testing prior to trials		
Stage 3: data analysis		
• Analysis of performance data against requirements	Micro	Implement
Full scale user trial	All	Implement/Deploy
Stage 3: data analysis		
• Capture expectations-reality gaps in terms of user experience of functionality through reflective interpretation of documentation analysis in the light of observations; through interviews and focus groups with teachers/students; and through critical incident analysis with students.	Micro	Implement/Deploy
• Capture expectations-reality gaps in terms of educational experience through reflective interpretation of documentation analysis and observations; through interviews/focus groups with teachers, students, museum educators; and through critical incident analysis with students.	Meso	Implement/Deploy
• Reflective analysis of expectations-reality gaps in terms of service impact	Macro	Deploy

Table 3: Methods used for data analysis during requirements and evaluation activities at each level, for each project phase

Furthermore, M3 provided an efficient way to structure data analysis for the evaluation of Myartspace, allowing the documentation of the potential of such a service. Successes and failures of the service at all levels (micro, meso and macro) were identified, along with inter-level influences. Table 3 outlines the data analysis processes for the data collected during stages 1 and 2 of the requirements and evaluation activities.

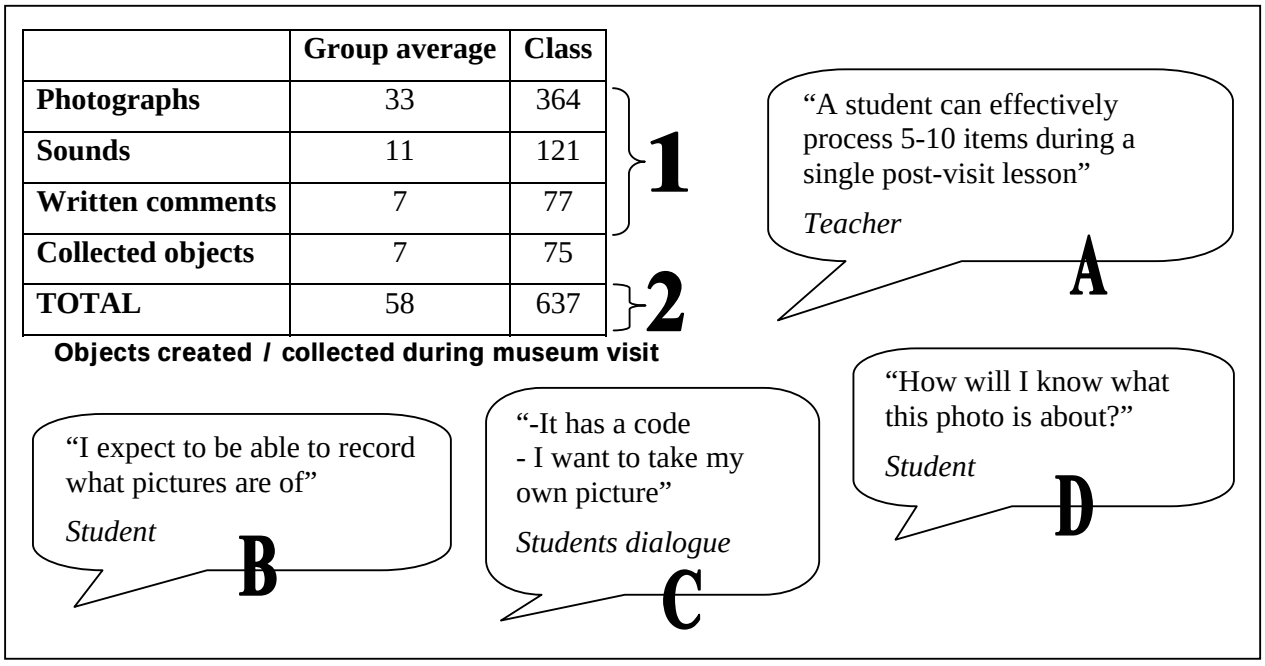

Figure 2: Data snippets gathered through observations (C, D), interviews (A, B), and system logs (1, 2)

Figure 2 gives an illustration of how M3 guided data analysis. It presents a sample of data snippets that were gathered through observations, interviews, and system logs. Snippet 1 shows that the children have created large amounts of content by taking photographs, recording sounds and writing comments. This suggests that at the micro level, creating and collecting items is a quick and easy task for them. Examination at the meso level reveals that this ease of use may not result in productive learning unless it is accompanied by creativity and sense of ownership, as exemplified in snippet C. This is an example of synergy between the two levels.

At the micro level, students are not able to annotate their photographs and the audio clips they recorded with notes describing what they were about or why they were recorded. Although they could create text notes, such notes could not be directly associated with photos or audio clips. This was puzzling for some students, who expected to be able to record this metadata (snippet B). Analysis at the meso level revealed frustration back in the classroom when students were trying to interpret their collections (snippet D). This is an example of a micro-level problem that migrates to the meso level, affecting the students' learning. Possible fixes to this problem can be placed either at the micro-level (e.g. changing the system to support annotation of photographs) or at the meso-level by giving advice to the students in effective techniques such as reading an exhibit label into the phone after taking a photograph of it, something that students were actually observed doing.

A final example comes from the analysis of snippets 2 and A. As mentioned previously, snippet 2 suggests that creating and collecting items is a quick and easy task at the micro level. Snippet A, however, suggests that decomposing and interpreting the collected content back in the classroom takes significantly longer, which resulted in students not managing to go through all their collected items during the post-visit lesson. This is an example of how omitting to resonate the micro and meso levels might lead to problems in the meso level. Possible fixes to this problem can be placed at any of the three levels. At the micro level, we can enforce an upper limit on the number of items a student can collect; or we might want to simplify the web interface so that online processing of items takes less time. At the meso level, we might try to educate students in regulating their collecting practices. At the macro level, we might try to influence the teachers' practice so that they include more than one post-visit lesson in their planning.

Discussion and Conclusion

M3 provides a structured format to assess usability, educational and organisational impact, and their inter-relationships. Table 4 describes how it follows the precepts for mobile learning evaluation presented earlier in the paper.

The six challenges in mobile learning evaluation identified in this paper are a direct consequence of the complex nature of mobile learning as we have come to understand it, as a social rather than technical phenomenon of people on the move constructing spontaneous learning contexts and advancing through everyday life by negotiating knowledge and meanings through interactions with settings, people and technology. In this paper we construed these challenges into precepts for evaluation and presented M3 as the implementation of one interpretation of these precepts; other frameworks previously proposed in the literature for the design (Ryu & Parsons 2008) and/or evaluation of mobile learning (Taylor 2004; Taylor *et al.* 2006; Traxler & Kukulska-Hulme 2005) can be seen as different interpretations of the same precepts. We view M3's main contributions to this growing body of knowledge and experience in mobile learning design and evaluation to be (a) its multi- and cross-level focus on individual interactions, educational processes and organisational change; (b) the way it combines with Lifecycle evaluation approaches to weave requirements analysis and evaluation into the whole development cycle, thereby emphasising experience-centred over technology-centred development; and (c) its focus on experience gains over cognitive gains alone. These qualities of M3 allude to Traxler's (forthcoming) proposal for alignment of our modernist conceptions of evaluation with the postmodern reality of mobile technologies and learning.

Precept	Framework qualities
P1. Capture and analyse learning in context; with consideration of learner privacy	• Illuminates learning activities and contexts at different levels of detail • Involves learners and teachers as informed participants in the evaluation process
P2. Assess the usability of the technology and how it affects the learning experience	• Micro-level (usability) evaluation activities are linked with evaluation activities at the meso and macro levels (educational effectiveness and institutional adoption) • The focus on interaction puts equal emphasis on the learners and the technology
P3. Look beyond cognitive gains into changes in the learning process and practice	• Relates the intended learning processes and outcomes to observed activities and examines the gaps between expectation and reality
P4. Consider organisational issues in the adoption of mobile learning practice and its integration with existing practices and understand how this integration affects attributes of in/formality	• Can analyse individual interactions, educational processes and organisational change • Can be applied across formal and informal settings
P5. Span the lifecycle of the mobile learning innovation that is evaluated, from conception to full deployment and beyond	• Integrates with a Lifecycle (Meek 2006) approach to evaluation

Table 4: M3 evaluation framework - fitness for purpose

The application of M3 in the context of Myartspace was successful and offered valuable insights to the project. Although we believe the framework is transferable to other mobile learning contexts, it needs further development to address, for example, contexts with higher ethical concerns, or contexts where it is challenging to align the requirements analysis and evaluation activities with the objectives and ethos of the project. The outcomes of an evaluation based on this framework can feed directly into system design, as has happened in the case of Myartspace. Perhaps with suitable extensions the framework could serve the design process more directly, guiding mobile learning designers to interpret and implement requirements for learning across self-constructed contexts.

Acknowledgements

We are grateful to Peter Lonsdale, Julia Meek and Paul Rudman for their invaluable contribution in the design of the evaluation of Myartspace and the development of the evaluation framework. We would like to thank the Department for Culture, Media and Sport for the funding of Myartspace through Culture Online. Myartspace was designed and developed by The Sea (http://www.the-sea.com); the service was branded as OOKL in 2007 and is now commercially available (www.ookl.org.uk) and used in a number of sites, including the Kew Gardens in London. Last, but certainly not least, we would like to thank all the students, teachers, and museum educators who took part in our trials, with special thanks to Bryony Kelly from the D-Day museum for her enthusiasm about Myartspace and her continuous support during the user trials.

References

Ahonen, M. (2008). Sustainable Wireless Computing – How Public Information Systems Can Be Designed to Reduce Exposure. In Proceedings of *31st IRIS Scandinavian Information Systems Conference*, Åre, Sweden,

Anastopoulou, S., Sharples, M., Wright, M., Martin, H., Ainsworth, S., Benford, S., Crook, C., Greenhalgh, C. & O'Malley, C. (2008). Learning 21st Century Science in Context with Mobile Technologies. In Traxler, J., Riordan, B. and Dennett, C., Proceedings of *mLearn 2008: The bridge from text to context*, Wolverhampton, UK, University of Wolverhampton, p. 12-19.

Bakardjieva, M., Feenberg, A. & Goldie, J. (2004). User-Centered Internet Research: The Ethical Challenge. In Buchanan, E.A. (Ed.), *Readings in Virtual Research Ethics: Issues and Controversies*, Idea Group Inc (IGI), p. 338-350.

Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, G., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J. & Thomas, D. (2001). *Manifesto for Agile Software Development*. Available online: www.agilemanifesto.org. Accessed on 22/01/09.

BECTA (2008). *Harnessing Technology: Next Generation Learning 2008-14*. Available online: http://publications.becta.org.uk/display.cfm?resID=38751&page=1835.

Black, P. & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice* **5**(1), p. 7-74.

Black, P. & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* **80**, p. 139-148.

Boud, D. (1995). Assessment and learning: contradictory or complementary? In Knight, P. (Ed.), *Assessment for Learning in Higher Education*. London, Kogan Page, p. 35-48.

Buchanan, E.A., (Ed.) (2004). *Readings in Virtual Research Ethics: Issues and Controversies*, Idea Group Inc (IGI).

Clough, G. & Jones, A. (2006). The Uses of PDAs and Smartphones in Informal learning. In Proceedings of *MLearn 2006*, Banff, Canada,

Colley, H., Hodkinson, P. & Malcom, J. (2003). *Informality and formality in learning*. Learning and Skills Research Centre. 1492/11/03/500.

Corlett, D., Sharples, M., Chan, T. & Bull, S. (2005). Evaluation of a Mobile Learning Organiser for University Students. *Journal of Computer Assisted Learning* **21**, p. 162-170.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin* **51**(4), p. 327-358.

Fountain, J.E. (2001). *Building the Virtual State: Information Technology and Institutional Change*, Brookings Institution Press.

Greenhalgh, C., French, A., Tennent, P., Humble, J. & Crabtree, A. (2007). From ReplayTool to Digital Replay System. In Proceedings of *e-Social Science Conference*, Ann Arbor, Michigan, USA,

Griffin, J. & Symington, D. (1998). Finding Evidence of Learning in Museum Settings. In Proceedings of *Evaluation and Visitor Research Special Interest Group Conference 'Visitors Centre Stage: Action for the Future'*, Canberra,

Hagen, P., Robertson, T., Kan, M. & Sadler, K. (2005). Emerging research methods for understanding mobile technology use. In Proceedings of *CHISIG*, Australia, p. 1-10.

Hidi, S. & Harackiewicz, J.M. (2000). Motivating the Academically Unmotivated: A Critical Issue for the 21st Centrury. *Review of Educational Research* **70**(2), p. 151-179.

Hsi, S. (2007). Conceptualizing Learning from the Everyday Activities of Digital Kids. *International Journal of Science Education* **29**(12), p. 1509 - 1529.

Kaplan, R.S. & Norton, D.P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, p. 75–85.

Kjeldskov, J. & Stage, J. (2004). New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies* **60**, p. 599-620.

Knight, P. (2001). *A Briefing on Key Concepts: Formative and summative, criterion and norm-referenced assessment*. LTSN Generic Centre.

Larman, C. & Basili, V.R. (2003). Iterative and Incremental Development: A Brief History. *Computer Supported Cooperative Work* **36**(6), p. 47-56.

Meek, J. (2006). *Adopting a Lifecycle Approach to the Evaluation of Computers and Information Technology*. PhD Thesis, School of Electronic, Electrical and Computer Engineering, The University of Birmingham.

Moore, R.W. & Sutman, F.X. (1970). The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching* **7**, p. 85–94.

Oliver, M. & Harvey, J. (2002). What does 'impact' mean in the evaluation of learning technology? *Educational Technology & Society* **5**(3), p. 18-26.

Pande, P.S., Neuman, R.P. & Cavanagh, R.R. (2000). *The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance*, McGraw-Hill.

Papadimitriou, I., Tselios, N. & Komis, V. (2007). Analysis of an informal mobile learning activity based on activity theory. In Vavoula, G.N., Kukulska-Hulme, A. and Pachler, N., Proceedings of *Workshop Research Methods in Informal and Mobile Learning*, WLE Centre, Institute of Education, London, UK, p. 25-28.

Patrick, K., Griswold, W., Raab, F. & Intille, S. (2008). Health and the Mobile Phone. *American Journal of Preventive Medicine* **35**(2), p. 177-181.

Price, S. & Oliver, M. (2007). A Framework for Conceptualising the Impact of Technology on Teaching and Learning. *Educational Technology & Society* **10**(1), p. 16-27.

Riley, D. (2007). Educational Technology and Practice: Types and Timescales of Change. *Educational Technology & Society* **10**, p. 85-93.

Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H. & Nyyssonen, T. (2004). *Examining mobile phone use in the wild with quasi-experimentation*. HIIT Technical Report 2004-1. Available online: www.hiit.fi/publications.

Royce, W.W. (1970). Managing the Development of Large Software Systems. In Proceedings of *IEEE WESCON*, p. 1-9.

Ryu, H. & Parsons, D. (2008). Designing Learning Activities with Mobile Technologies. In Ryu, H. and Parsons, D. (Eds.), *Innovative Mobile Learning: Techniques and Technologies*, Idea Group Inc (IGI), p. 1-20.

Scriven, M. (1967). The methodology of evaluation. In Tyler, R.W., Gagne, R.M. and Scriven, M. (Eds.), *Perspectives of Curriculum Evaluation*. Chicago, Rand McNally, p. 39-83.

Sharples, M., Taylor, J. & Vavoula, G. (2007a). A Theory of Learning for the Mobile Age. In Andrews, R. and Haythornthwaite, C. (Eds.), *The Sage Handbook of E-learning Research*. London, Sage, p. 221-47.

Sharples, M., Lonsdale, P., Meek, J., Rudman, P.D. & Vavoula, G.N. (2007b). An Evaluation of MyArtSpace: a Mobile Learning Service for School Museum Trips. In Proceedings of *mLearn 2007*, Melbourne, Australia,

Sharples, M., Jeffery, N., Du Boulay, J.B.H., Teather, D., Teather, B. & Du Boulay, G.H. (2002). Socio-cognitive engineering: a methodology for the design of human-centred technology. *European Journal of Operational Research* **136**(2), p. 310-323.

Smith, H., Heldt, S., Fitzpatrick, G., Hui Ng, K., Benford, S., Wyeth, P., Walker, K., Underwood, J., Luckin, R. & Good, J. (2007). Reconstructing an informal mobile learning experience with multiple data streams. In Vavoula, G.N., Kukulska-Hulme, A. and Pachler, N., Proceedings of *Workshop Research Methods in Informal and Mobile Learning*, WLE Centre, Institute of Education, London, UK., p. 29-34.

Taylor, J. (2004). A task-centred approach to evaluating a mobile learning environment for pedagogical soundness. In Proceedings of *MLearn2004*, London, UK, Learning and Skills Development Agency, p. 167-171.

Taylor, J., Sharples, M., O'Malley, C., Vavoula, G. & Waycott, J. (2006). Towards a Task Model for Mobile Learning: a Dialectical Approach. *International Journal of Learning Technology* **2**(2/3), p. 138-158.

Traxler, J. (forthcoming). Mobile Learning Evaluation: The Challenge of Mobile Societies. In Vavoula, G., Pachler, N. and Kukulska-Hulme, A. (Eds.), *Researching Mobile Learning: Frameworks, Methods and Research Designs*. Oxford, Peter Lang,

Traxler, J. & Bridges, N. (2004). Mobile Learning - The Ethical and Legal Challenges. In Proceedings of *MLEARN 2004*, Bracciano, Italy,

Traxler, J. & Kukulska-Hulme, A. (2005). Evaluating Mobile Learning: Reflections on Current Practice. In Proceedings of *MLEARN 2005*, Cape Town, South Africa,

Trinder, J., Roy, S. & Magill, J. (2007). Have You Got Your PDA With You?...Denials and Accusations. In Vavoula, G.N., Kukulska-Hulme, A., & Pachler, N., Proceedings of *Workshop Research Methods in Informal and Mobile Learning*, Institute of Education, London, UK,

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

Vavoula, G., Sharples, M., Rudman, P., Lonsdale, P. & Meek, J. (2007). Learning Bridges: a role for mobile technologies in education. *Educational Technology & Society* **XLVII**(3), p. 33-37.

Vavoula, G., Sharples, M., Rudman, P., Meek, J. & Lonsdale, P. (forthcoming). Myartspace: Design and evaluation of support for learning with multimedia phones between classrooms and museums. *Computers and Education*.

Vavoula, G., Sharples, M., Rudman, P., Meek, J., Lonsdale, P. & Philips, D. (2006b). Museum explorations in physical, personal and virtual space through MyArtSpace. In Proceedings of *mLearn 2006*, Banff, Canada,

Vavoula, G.N. (2005). *WP4: A Study of Mobile Learning Practices*. MOBIlearn deliverable D4.4. Available online: http://www.mobilearn.org/download/results/public_deliverables/MOBIlearn_D4.4_Final.pdf