

Evaluating Mobile Learning: Reflections on Current Practice

John Traxler, Wolverhampton University, Centre for Learning and Teaching, United Kingdom, *John.Traxler@wlv.ac.uk*

Agnes Kukulska-Hulme, The Open University, Institute of Educational Technology, United Kingdom, *a.m.kukulska-hulme@open.ac.uk*

Abstract:

The field of mobile learning is at present characterised by a proliferation of pilots and trials that allow mobile technologies to be tested out in a variety of learning contexts. The sustained deployment of mobile learning will depend on the quality of these pilots and trials, which includes evaluation methodology and reporting. The paper examines current evaluation practice, based on evidence drawn from conference publications, published case studies, and other accounts from the literature. The authors also draw on their work in collecting case studies of mobile learning from a range of recent projects. Issues deserving examination include the apparent objectives of the pilots or trials, the nature of the evaluations, instruments and techniques used, and the analysis and presentation of findings. The paper reflects on the quality of evaluation in mobile learning pilots and trials, in the broader context of evolving practices in the evaluation of educational technologies.

Keywords:

Evaluation, methodology, literature review, good practice.

1. Introduction

The last four years have seen a dramatic increase in the number and the variety of pilots and trials involving mobile learning. This increase has been reflected in and supported by submissions to the MLEARN conference series, and by the emergence of smaller and local conferences. The growing pedagogic and technological sophistication of mobile learning pilots and trials is evident, but increased and sustained deployment of mobile learning will depend of the quality of analysis and evaluation of these pilots and trials.

Evaluation and analysis are key to the sustainability of mobile learning because first, they inform the outside world about the effectiveness of pilots and trials, specifically in relation to the objectives they set out to achieve, and second, they provide some insights to funders and champions on the utility and cost-effectiveness of the pilots and trials they support. In some cases, evaluation will determine whether pilot projects become embedded as ongoing provision in institutions.

Historically, the development of evaluation strategies for educational trials and pilots has focused on face-to-face contact with students or learners working in cohorts on courses in classrooms, lecture theatres and laboratories, and perhaps involving other stakeholders. Evaluation has depended on a small stable repertoire of techniques, embracing, usually, observation, interview, focus group and questionnaire. More recently, e-learning evaluation has supplemented these with analysis of system logs and technical data. In other areas of social research, for example policy interventions, and to a lesser extent, in other areas of educational research, evaluation is a complex and mature discipline (Patton 2001).

Now, the changing political, economic and social climate is forcing or encouraging educational institutions

- to address new constituencies of learners, such as ‘access’ students without adequate study skills and full-time students forced to hold down substantial part-time jobs
- to deliver informal and life-long learning, alongside conventional structured courses and programmes
- to engage with industry and commerce by delivering more training
- to teach increasing numbers of students in spite of static financial resources.

This means that developmental and innovative trials and pilots working in these environments, in our case those that use mobile technologies, must question the efficacy of traditional evaluation techniques and must adapt and explore more innovative and varied evaluation approaches. This is in itself

problematic because it raises concerns about the validation of untried evaluation techniques and about the trustworthiness of adapted techniques.

There are no *a priori* attributes of a 'good' evaluation (to say that there would be to take an implicitly realist or essentialist position that not every stakeholder would agree with, and would also confront a widely held view that in fact evaluation is a contingent activity). In an earlier work, we tried to outline some candidate attributes of a 'good' evaluation (Traxler 2002) but we also identified the reasons why evaluation of mobile learning is unusually challenging. Briefly some of these attributes were that a 'good' evaluation should be:

- Rigorous, meaning that conclusions must be trustworthy and transferable
- Efficient, in terms of cost, effort, time
- Ethical, specifically in relation to the nuances of evolving forms of provision
- Proportionate, that is, not more ponderous, onerous or time-consuming than the learning experience or the delivery and implementation of the pilots themselves
- Appropriate to the specific learning technologies, to the learners and to the ethos of the project concerned – ideally *built in*, not *bolted on*
- Consistent with the teaching and learning philosophy and conceptions of teaching and learning of all the participants
- Authentic, in accessing what learners (and perhaps teachers and other stakeholders) *really* mean, *really* feel, and sensitive to the learners' personalities within those media
- Aligned to the chosen medium and technology of learning
- Consistent across:
 - different groups or cohorts of learners in order to provide generality
 - time, that is, the evaluation is reliably repeatable
 - whatever varied devices and technologies are used

The last of these attributes is particularly challenging in mobile learning, since the technologies are changing at an exceptionally fast pace and consequently reaching an understanding of underlying issues is difficult. Some of the others are more subtle. Some issues around ethics have been explored elsewhere recently (Traxler and Bridges 2004): mobile learning continues to evolve however, and any account of the ethics of evaluation would now, for example, have to address the issue of using virtual ethnography to evaluate mobile blogging.

As we have said, there is no *a priori* basis for these attributes and thus no specific philosophical basis for the agenda of this paper, namely reflection on current evaluation practice in mobile learning trials and pilots. However we feel that in light of the current state of evaluation in mobile learning, we can still make a valuable contribution to theory and practice by juxtaposing some current evaluation practice as illuminated by the literature with some discussion of 'good' evaluation.

Bates and Poole (2003) have proposed a model for the effective use of technology for teaching in higher education that suggests eight criteria to be used in determining choice of technology. An investigation of whether the right technology has been selected is arguably an important aspect of a comprehensive evaluation of mobile learning. It would therefore have to take account of these criteria, namely:

- the appropriateness of the technology for students
- ease of use and reliability
- costs
- teaching and learning approaches
- interactivity
- organizational issues
- novelty, as a choice not to use existing technology
- speed, i.e. how quickly materials can be developed

It is interesting to note that this model is an evolution of a previous model developed by Bates (1988, 1995), which had 'access' as one of its key criteria - in the sense of providing flexibility, or reaching students who could not attend conventional classes. This criterion was subsequently de-emphasized and subsumed under 'appropriateness for students'. However, our own analysis of the reasons why teachers are using mobile technologies (reported below) suggests a possible reinstatement of access as a key criterion in relation to these new technologies.

There are a wide variety of other authors offering complementary and perhaps competing criteria for all the various aspects of evaluating mobile learning; one way of addressing this complexity and of

exploiting growing experience and expertise would be to use (online) Delphi techniques to develop more consensual criteria (see for example, Des Marchais 1999), and perhaps to use ‘contrived’ elicitation techniques (Rugg and McGeorge 1992) to uncover the value systems that underlie experts’ evaluation criteria.

We can make some progress on the basis that mobile learning pilots and trials each have their own aims and objectives, and that these have driven evaluation in the sense of defining the outcomes sought by the evaluation and hence driving the selection and development of the techniques, instruments and protocols used in evaluation. The outcomes we are interested in are broadly ‘educational’. A different evaluation mindset may apply to a project that sets out to test only technical stability, or to trial a specific interface design without explicitly addressing its impact on an educational activity.

Perhaps the obvious potential objective for any educational pilot or trial is cognitive change, where students have learnt something new. An evaluation may be looking for this or might be looking for meta-cognitive change, where students have learned something about the process of learning. An evaluation may also be looking for affective changes in students, reflecting changed feelings, values or preferences and it may also look for social changes, perhaps in how students relate to or work with each other, or in how groups of students show increased collective interaction, competences or skills.

Mobile learning takes place in a wider social context and evaluation must also recognise this. The wider social and economic benefits of projects may be evaluated through the eyes of learners and other stakeholders, if the pilot or trial has been funded with a social and economic agenda. There is increasing recognition of such benefits but also the difficulty of evaluating them appropriately (Dewson *et al.* 2002). Mobile learning pilots and trials can nowadays receive funding to build capacity within communities and to improve social inclusion; the pan-European *m-learning* project is an obvious example of such an agenda, as are its smaller spin-offs with travellers and the homeless. In these cases, lifelong learning and community education might only be the means to these ends and mobile learning technologies may only be the vehicles to carry them. In such cases, the cognitive gains are likely to be irrelevant – indeed, any kind of learning gains might be irrelevant, nevertheless evaluation must still take place.

2. Current Evaluation Practice

This section reviews current mobile evaluation practice, as described in recent reported case studies, pilots and trials (Kukulka-Hulme and Traxler 2005; Attewell and Savill-Smith 2004). The sample is relatively small and arbitrary, and of course the accounts themselves may be an incomplete or inaccurate reflection of the actuality. The case studies (Kukulka-Hulme and Traxler 2005) were written to a template specifically designed to produce consistent accounts that exposed the relationships between objectives, methods and evaluation; the conference contributions (Attewell and Savill-Smith 2004) were often shorter and less consistent and may often have discussed technologies that were not mature enough for use across a cohort of learners within a formal course for a semester or a term of their studies.

The issues and questions addressed in this section are drawn from a much larger developing evaluation framework being developed by the authors that attempts to build systematically and comprehensively on secure philosophical foundations. A motivation behind the current exercise has been to look at priorities and gaps in this framework, and to create dialogue with practitioners on the issues raised.

In general, examining most accounts of trials and pilots produces only indicative rather than definitive findings but these nevertheless serve to help us develop a more focused and systematic account of evaluation. The following sections briefly highlight how the multiplicity of aims and objectives pose challenges for evaluation; look at some of the evaluation techniques and methods used and at how findings are presented.

2.1 Aims and Objectives of Pilots and Trials

An analysis of 12 international case studies in Kukulka-Hulme & Traxler (2005) reveals that reasons given for using mobile technologies in teaching and learning relate principally to improving access, exploring changes in teaching and learning, and alignment with institutional or business aims, as illustrated by these examples:

Access:

- Improving access to assessment, learning materials and learning resources

- Increasing flexibility of learning for students
 - Compliance with special educational needs and disability legislation
- Changes in teaching and learning:
- Exploring the potential for collaborative learning, for increasing students' appreciation of their own learning process, and for consolidation of learning
 - Guiding students to see a subject differently than they would have done without the use of mobile devices
 - Identifying learners' needs for just-in-time knowledge
 - Exploring whether the time and task management facilities of mobile devices can help students to manage their studies
 - Reducing cultural and communication barriers between staff and students by using channels that students like
 - Wanting to know how wireless/mobile technology alters attitudes, patterns of study, and communication activity among students
- Alignment with institutional or business aims:
- Making wireless, mobile, interactive learning available to all students without incurring the expense of costly hardware
 - Delivering communications, information and training to large numbers of people regardless of their location
 - Blending mobile technologies into e-learning infrastructures to improve interactivity and connectivity for the learner
 - Harnessing the existing proliferation of mobile phone services and their many users.

A review of the 27 projects documented in the proceedings of MLEARN 2003 (Attewell and Savill-Smith 2004) shows a similar spread of objectives, with a predominance of objectives identifying or targeting changes in teaching and learning:

Access:

- enabling students to look at course information any time and anywhere
- trying to ensure that every student can access content independently of the channel he or she chooses to use
- the use of a PDA as an assistive technology
- ensuring that classroom-based pupils benefit from the experience of a field trip being undertaken by their peers

Changes in teaching and learning:

- individualisation:
 - to explore the potential for individualised mobile learning - revision material tailored to the needs of the individual
 - to provide learners with a flexible context-awareness system that can react to their needs
- collaborative and active learning:
 - immediate feedback through interactive tests: the user knows in real time if their choice is correct
 - interactive screens encouraging art gallery visitors to respond to the art on view
 - a set of innovative games, materials and activities which will motivate reluctant young learners
 - a user-friendly m-portal that is powerful and empowering, and encourages active participation by its users
 - enhancing interactivity and cooperation while preserving the traditional advantages of face-to-face encounters
- informal learning with multiple media:
 - to investigate how self-produced videos, made with a digital video camera and later viewed on handheld mobile computers, can support informal learning
 - to provide video and still images giving additional context for art gallery works on display, opportunities to listen to an expert talk about details of a work, with the details simultaneously highlighted on the screen
 - enhancing the audio presentation of a multimedia museum guide by using the PDA screen to travel throughout a fresco and identify the various details in it
 - using voice technology to provide rich media content for the user
- cognitive and behavioural change:
 - to explore how context-dependent learners' knowledge concepts are

- to evaluate fragmentation in mobile learning based on students' deep and surface approaches to learning
- to capture learners' thoughts, views and behaviours in a mobile learning setting

Alignment with institutional or business aims:

- to remain at the cutting edge of educational technology by helping to shape a new generation of multimedia tours in art galleries
- to investigate whether an integrated set of learning tools would be useful, which tools would be adopted and the contexts in which the tools would be used
- development of a service model and new component concepts for lifelong mobile learning

In the MLEARN 2003 proceedings we have also identified some aims that address 'the future of mobile learning' more generally:

- to find out in which arenas handhelds are used, how and why they are used, and what role they can play
- what the future take-up of new services and facilities on mobile phones and other technology devices might be
- to find out whether young adults would be willing to use their phones for literacy and numeracy learning
- to understand the range of actions and opportunities open to mobile learners, and seek ways of extending this range to support what learners want to do – even if they themselves do not yet know what that is

The authors of the case studies and papers do not always categorise their own objectives, and they do not prioritise them explicitly. Objectives can usually be identified quite clearly in the abstracts, whilst in the actual papers they may be introduced and discussed at various points in the paper. They are not presented as being linked to evaluation methods in any explicit way, even when the methods are fully explained and justified. Objectives typically consist of the provision of new opportunities, exploration of potential, introduction of new media, or initial exploration of attitudes and patterns of use. The authors may be seeking to encourage or enhance certain ways of learning, but such outcomes may be hard to measure, especially during a relatively short period of technology use.

Given the present state of knowledge in the field of mobile learning, considerable skill is required in formulating objectives. In research projects, the aims may be exploratory and may seek to retain a certain openness to unexpected findings. Clearly, this is problematic for some styles of formal summative evaluation that audit achievements against objectives without recognising that some objectives in exploratory projects may only emerge as projects evolve (whilst it may also become apparent that some initial objectives are disproportionately expensive or impractical to realise). A more fluid, formative and engaged style of evaluation may on the other hand run the risk of losing its perspectives or objectivity.

These are some of the issues that we are addressing in exploring how mobile learning evaluation should engage with project objectives; they are issues with a very direct bearing on the methods and techniques of evaluation.

2.2 Types of Evaluation

In looking at our sample, we find that most of the projects can be characterised as 'first-generation' – the technologies used were not always stable, mature or well understood and technical difficulties sometimes hampered the educational delivery and the subsequent evaluation. The sample also seemed to be generally 'first generation' in the sense that most of the trials and pilots were not building on an established base of expertise or equipment. It could be argued that this had consequences for the type of objective that was appropriate, shifting the balance away from answering specific research questions and towards identifying, refining or prioritising such questions (or perhaps demonstrating technological and pedagogic possibilities). Whichever it was, this issue has considerable implications for evaluation, both in the general sense of understanding the technological and pedagogic contexts but also in the specific sense of evaluating the alignment of objectives, methodology and outcomes.

The evaluations in our sample were usually formative (in the sense that changes took place as a consequence). It was not always apparent why they took place when they did. Insofar as we could judge, the evaluations were not usually conducted by external staff or by evaluation specialists.

Very few of the accounts in our sample made direct reference to the literature of evaluation and consequently there was not normally a justification or explanation for the evaluation methods or instruments chosen. There were also few, if any, references to ethics as a potential factor in evaluation in the literature reviewed. The exceptions related to privacy and context-awareness (Lonsdale *et al.*, 2004).

2.3 Methods and Techniques

The vast majority of pilots and trials in our sample had no explicit or apparent educational or epistemological foundations. They may nevertheless have had some tacit foundations that were ‘not-worth-mentioning’ or ‘taken-for-granted’. Where foundations were apparent, they were usually social constructivist and underpinned the pedagogy. There was not usually any statement about the foundations of the evaluation. The fact that foundations were not articulated was problematic because basing a pilot or trial on explicit educational or epistemological foundations implicitly raises the issue of whether the evaluation should be aligned with them.

A crucial element of the evaluation of mobile learning trials and pilots is the elicitation of learners’ attitudes and achievements. In examining elicitation methods and techniques used in each one, the issues can be categorised as:

- Technical: were the methods and techniques used competently and were they appropriate to the specific context of the project?
- Philosophical: what was the relationship between the evaluation elicitation methods and the project in terms of their actual, assumed or implied epistemologies?
- Ethical: were the elicitation methods used with an ethical basis that was consistent with actual, assumed or implied epistemologies?

The technical questions were perhaps the easiest to answer since most authors address them explicitly though not always in much detail. The elicitation techniques used were usually questionnaires, interviews and focus groups, supplemented by observation of the learners or analysis of system data.

In general where these techniques were used, their implementation and delivery was conventional rather than adapted to mobile technologies or mobile learning. If the pilot or trial being evaluated took place in the context of blended learning or classroom delivery this would be appropriate but where the trial or project focussed purely on informal mobile learning, it could be argued to be methodologically flawed and possibly ethically problematic. In fact, most accounts of evaluations did not extend to a discussion of selecting and designing the instruments used.

Questionnaires usually used Likert scale fixed responses and the analysis took the form of bar-charts or histograms. The samples and populations were too small for any form of statistical analysis, though the issue of confidence was usually underplayed and results were frequently translated into percentages or fractions. The fixed response questions were often supplemented by free-text answers. Their analysis took the form of paraphrase.

Interviews, focus groups and observations were used less frequently. Accounts of interviews and focus groups were brief and suggested that the sessions were usually short and probably informal. In a relative small number of cases, evaluation used and combined several of these different techniques.

2.3 How findings are being presented

Apart from the usual academic reporting of findings, an evaluation may result in recommendations. Attewell and Savill-Smith (2004) include a number of papers that make recommendations based on evaluation findings: Mitchell gives recommendations and key challenges for developers; Vainio and Ahonen discuss their work in the MOBIlearn project in terms of providing guidelines for adaptive user interface design; Thomas *et al.* put forward design principles for the development and evaluation of mobile educational games, but emphasize their tentative quality:

“While the principles presented here provide a conceptual overview of what could become ‘good practice’ in relation to the development and evaluation of mobile learning games, only a handful of principles have been illustrated... The aim of this paper has been to highlight the need for comprehensive research and evaluation of usability principles and initiate a dialogue in which usability frameworks can be modified, enhanced and validated by the mobile learning community.”

(Thomas *et al.* in Attewell and Savill-Smith 2004, p.179)

The multiplicity of stakeholders in mobile learning pilots and trials and the fluidity of terminology and technology make the presentation of findings (but also the presentation of evaluation) an important issue. One role for evaluation specialists, especially where external funding or public money is involved, is placing findings, outcomes and achievements in some wider context.

3. Conclusions

To draw some of the threads of our discussion together, we have said that developing the concept of a 'good' evaluation is problematic but important. Such a concept should be based on explicit philosophical foundations but our review shows this is not usually the case in mobile learning evaluation at the moment. The evaluations we examined seem to be based on tacit foundations of 'common-sense' or perhaps a tacit consensus amongst mobile learning evaluators. It seems to us that many of the trials and pilots themselves rest on a 'common sense' view of learning. In the case of evaluation, this means that there is not always much theoretical justification or coherence to support the selection or use of any given evaluation techniques or methods. These facts jeopardise the credibility of outcomes reported in mobile learning trials and pilots. Mobile learning challenges evaluators to develop evaluation methods and techniques that are sympathetic to the ethos and technologies of mobile learning but our review shows that these are slow to emerge.

Recommendations for good practice can only convincingly rest on evaluation and consequently recommendations for good evaluation can only rest convincingly on an evaluation of evaluations. We hope we have contributed to that process but recognise the need for continued work on a more systematic and comprehensive framework. Evaluation is problematic but valuable. In the words of one experienced educational evaluator (Somekh 2001):

“Evaluation is a fascinating, socially useful, morally demanding and highly politicised activity. Its future depends on the uses we put it to, and the role it is given by sponsors and politicians.”

References

- Attewell J and Savill-Smith C (2004). *Learning with mobile devices - research and development - a book of papers*. Learning Skills Development Agency: London.
- Bates AW (1988). Technology for distance education: A 10 years' perspective. *Open Learning*, 3(3).
- Bates AW (1995). *Technology, Open learning and Distance Education*. London: Routledge.
- Bates AW and Poole G (2003). *Effective Teaching with Technology in Higher Education*. San Francisco: Jossey-Bass.
- Des Marchais JE (1999). A Delphi Technique To Identify And Evaluate Criteria For Construction Of PBL Problems, *Medical Education* 33(7) 504 – 508.
- Dewson S, Eccles J, Tackey N D and Jackson A (2002). *Measuring Soft Outcomes and Distance Travelled: A Review of Current Practice*. Institute for Employment Studies for DfEE: London.
- Kukulka-Hulme A and Traxler J (2005) *Mobile Learning: A Handbook for Educators and Trainers*. Routledge: London.
- Lonsdale P, Baber C, Sharples M and Arvanitis T N (2004). A context-awareness architecture for facilitating mobile learning, in Attewell J, Savill-Smith C (Eds) *Learning with Mobile Devices: Research and Development*, London: Learning and Skills Development Agency.
- Patton M Q (2001). *Qualitative Research and Evaluation Methods*. SAGE Publications: London.
- Rugg G and McGeorge P (1992). Uses of Contrived Knowledge Elicitation Techniques. *Expert Systems*, 149-154.
- Somekh B (2001). The Role of Evaluation in Ensuring Excellence in Communications and Information Technology Initiatives. *Education, Communications and Information*, 1, 75-101.
- Traxler J (2002). Evaluating m-learning. *Proceedings of MLEARN 2002*, European Workshop on Mobile and Contextual Learning, 63-64, University of Birmingham, 20-21 June 2002.
- Traxler J and Bridges N (2004). Mobile Learning - The Ethical and Legal Challenges, *Proceedings of MLEARN 2004*, Bracciano, Italy, June 2004.