
A toolkit for supporting evaluation

Gráinne Conole,* Ed Crew,* Martin Oliver** and Jen Harvey***

*University of Bristol, **University College London

***Dublin Institute of Technology

email: g.conole@bristol.ac.uk

This paper describes the development of a Web-based evaluation toolkit that supports practitioners in the effective and appropriate use and evaluation of learning materials. An outline of the basic toolkit architecture and design will be described, along with extracts of some of the formative feedback received on the use and value of the toolkit during a range of user trials which were carried out as part of the project development work. The paper will conclude with potential applications for such toolkits and recommendations for future development.

Introduction

Despite the current consensus that adapting and reusing existing learning resources, including Information and Communication Technology (ICT), is a good thing, examples of this practice are few and far between. This can be traced to a number of factors. In particular, the 'not invented here' syndrome (HEFCE, 1996) is no doubt still present. However, more important is the issue of the time and skills required to retrieve, evaluate and then adapt materials. This is compounded by the fact that identifying suitable resources in the first place can be complex (although the growth of subject-specific information gateways, portals and guidelines to resources will go some way towards alleviating this problem). Some of the barriers to uptake that have been identified include:

- the quantity of material, of varying quality, now available through the Web;
- the problem of retrieving appropriate materials;
- the difficulty of adapting other peoples' materials;
- the issue of ownership and copyright;
- integration of new materials, including issues of style, definition and level into existing courses;

- staff having the appropriate educational and/or technical skills required to evaluate, adapt and integrate materials;
- concern about the currency of materials, particularly accuracy and whether it is up to date.

Furthermore, few academic staff have had the opportunity to develop the prerequisite expertise to design and implement an effective strategy for acquisition, use and evaluation of either new materials or innovative methods of delivery. In order to provide these practitioners with support and encouragement, easy-to-use guidelines and resources are required. Not only must these be based on sound, tested pedagogic theories, but first and foremost they must be practical for academics using (or evaluating the use of) learning technologies. This paper will describe an evaluation toolkit that addresses these concerns by providing simple step-by-step guidance on the development of effective evaluation strategies for learning and teaching.

Background: providing support and guidance for decision-making

Resources for supporting decision-making range from highly restrictive ‘templates’ or ‘wizards’, which provide step-by-step guidance but little possibility of user-adaptation, through to ‘theoretical frameworks’, which provide a context and scope for the work but leave the user to devise their own strategy for implementation. Between these extremes lies a range of resources, including checklists, guidelines and step-by-step tutorials. This project set out to develop a type of resource which lies approximately half way along this continuum, referred to here as a ‘toolkit’.

Toolkits include an expert model of a system derived from recognized theory and best practice. This is used to provide a structured process, supporting the implementation of performance monitoring systems. Furthermore, by providing a common conceptual framework (particularly one in which multiple interpretations of terms can be negotiated and agreed), it becomes possible to define and establish standards.

A more detailed description and definition of frameworks, toolkits and wizards, supported by illustrative examples, is provided elsewhere (Conole, Oliver, and Harvey, 2000).

An outline of the toolkit

The ‘Evaluation Toolkit for Practitioners’ project is funded by JISC as part of the Committee for Awareness, Liaison and Training’s remit. It has developed and tested a Web-based toolkit that helps practitioners, irrespective of their current degree of expertise, to evaluate their selection and use of learning materials. It provides a structured resource that can be used to plan, scope and cost an evaluation. By providing progressively more detailed information on particular topics, the user can follow up relevant issues when and if this is required. Secondly, by providing a simple, logically organized structure the toolkit reduces the time required to plan work of this type. The aim is that the toolkit will be used iteratively, with progressively more detailed analysis occurring once initial feedback and information has been received.

The development of the toolkit was built on a set of underlying assumptions. It should:

- be easy to use for practitioners and provide demonstrable benefit;

- integrate guidance, but not be prescriptive;
- be adaptable and easy to customize to the local context;
- enable access to a comprehensive resource of relevant material.

The toolkit in its current format is designed to be introduced initially as part of a structured face-to-face workshop session. Users can register for a username and password and are then given access to their own working area of the toolkit. They can then build up a portfolio of different evaluations to suit their needs. Each user also has access to a set of predefined case study evaluations that they can use for guidance or as a model that can be customized to create their own evaluation plan.

The toolkit is run online from a server at the University of Bristol (<http://www.ltss.bris.ac.uk/jcalt/>). Users have access to their own private workspace within the toolkit, and can store any number of evaluations within this area. Summaries of each evaluation can be printed out in a number of formats, either by evaluation topic or as a summary of the whole evaluation. The toolkit is an SQL-database-driven Web application, which provides a tailored interface to a large resource of content and external links. The basic framework consists of over 200 files, which include scripts to generate appropriate HTML pages of content and resources. It is designed so that output depends on user input with results being displayed 'on the fly' in a user-orientated manner. The resource includes embedded links to a large range of relevant external resources as well as a dynamic database of toolkit resources. The database of resources can be adapted and is designed to be expandable to include new resources or information as required. In addition, the toolkit is designed so that users can adapt and define personal requirements, categories or content. The resource also includes a feature that allows users to share content through contribution of evaluation case studies to a general pool of resources.

Toolkit sections

The toolkit consists of three components: i) Evaluation Planner, ii) Evaluation Advisor and iii) Evaluation Presenter. There are also links to three databases on data capture methods, data analysis and presentation techniques. A key feature of the resource is that it employs a filtration process to recommend relevant materials to the user at appropriate stages of the evaluation. Each section also contains substantial links to related material, such as *The LTDI Evaluation Cookbook* (Harvey, 1998) and other Web resources.

Although the Toolkit is presented in three sequenced sections, the modular format of the structure is designed to enable use of the different components when and if required during an ongoing evaluation study. This means that a user can access their personal workspace and retrieve, amend and develop their evaluation as they are carrying out the work rather than having to complete all the sections before undertaking the work.

Evaluation Planner

The first section of the toolkit helps users to define the scope of their evaluation and is divided into seven main stages. The content is linked with a series of questions, guidelines and exercises supported by user input fields, which guide users in planning their evaluation strategy and clarifying their reasons for particular selections.

The section also supports users in the identification of the intended audience for the

evaluation (stakeholder analysis) and the definition of the core evaluation question(s). It focuses on defining the complexity or scale of the proposed evaluation, giving particular attention to associated time and resource implications (both financial and human). Users are expected to give some thought at this stage to defining evaluation tasks, roles and responsibilities. The output from Planner includes an evaluation strategy and implementation guide. This provides the starting-point for the communication plan in the Evaluation Presenter section of the toolkit.

After the introductory 'About Planner' stage, the users are asked to specify the focus of the evaluation in the 'What are you evaluating' section. The user is given a predefined list of potential evaluation objects to choose from (Figure 1). They are also able to add or edit their own entries.

The 16 evaluation objects you have to choose from

evaluation object	explanation	
Course materials	Could be self-contained static resources	1
	Evaluation might focus on either design or use	2
Course programmes	Could be a complete programme of study	3
	Might be at module or course level	4
	Could involve evaluation of all aspects of the learning experience	5
	Might be a longitudinal study	6
External resources	Could be to assess whether or not they are appropriate to be included in a course	7
	Could include some evaluation of the associated costs and benefits	8
Externally-funded project	This could be either a developmental or research project	9
	The evaluation might be an assessment of the success of the project or might be more formative, with the intention being to improve the project through the evaluation results.	10
New innovation in learning and teaching	This might be where you have replaced or supplemented your teaching with some use of ICT or might be where you have enhanced the course by moving from individual to group-based learning	11

Figure 1: Suggested evaluation objects.

This process of guiding the user through a set of choices, whilst maintaining user freedom by allowing them to customize or add to the entries, is an integral feature throughout the toolkit. In a similar fashion, the user is guided through the process of thinking about their reasons for carrying out the evaluation (and any associated resource implications) and are asked to describe the context for the study. These sections consist of a mixture of user-input boxes and choice lists (Figure 2).

The user-input boxes are associated with questions designed to prompt reflection on the evaluation design process. At any stage during the evaluation, the user can return to earlier steps and edit entries, to support this reflective approach and encourage iterative development of the evaluation plan.

1 Resources	<p>Is there a group of learners or users associated with this?</p> <p>If there are learners or users associated with the study, give details of the number and their level/background experience.</p> <p>Who will be doing the evaluation?</p> <p>What resources have you got available?</p> <p>Is there a budget to support the evaluation activities?</p>	<p>The project is aimed at lecturers in HE with little or no experience of evaluation. The evaluation will be</p>
2 Scale	<p>Include here details of the size or scope of what you are evaluating, for example is it two web pages or 400?</p> <p>If relevant, how many learner hours does it represent?</p> <p>Is it self contained or integrated with other learning activities?</p>	<p>The time required to use the toolkit needs to be established, but we estimate that it will take several</p>
3 Timeframe	<p>What will this enable you to do</p> <p>When are you planning to undertake this work and when does it need to be completed?</p>	<p>The work needs to be squeezed in alongside the development activities, within a short (6 month)</p>

Figure 2: Examples of questions prompting reflection on the context of the study.

Lecturer - Timesaving			
The stakeholder and their concerns >	Lecturer	Lecturer	Student
	Timesaving	Research	Usability
Exploratory questions Thinks about a question with a stem using words and phrases such as, 'what', 'when', 'for whom', 'under which circumstances', and so on.	You may want to start your question with one of these stems: <input type="text"/>		
	<input type="text"/>		
	<input type="text"/> no relevance <input type="text"/> Relevance to Lecturer - Timesaving		
Comparative questions Consider questions looking at alternative	You may want to use one of the comparisons below: <input type="text"/>		
	<input type="text"/>		

Figure 3: Suggestions of different ways to rephrase the evaluation question.

Particularly important in this process are the section summaries and the overall final evaluation plan. Once the user has clarified the nature of the evaluation and the reasons for doing it, they are asked to look at the stakeholders involved in the process. As before, the user is given the option of a predefined set of potential stakeholders and of the concerns that each of these might wish to have addressed. Users are encouraged to limit the number of stakeholders and concerns to a maximum of three to ensure that a management evaluation plan can be devised.

The next step guides the user through the derivation of evaluation questions to meet the individual stakeholder concerns. Previous research has shown that practitioners find this step surprisingly difficult (Oliver and Conole, 2000) and that a structured approach to deriving the question can support them in this process. The user works through a number of alternative ways of thinking about each stakeholder concern, guided by a set of question types and potential question stems (Figure 3).

Having worked through this stage the user defines the final questions for each of the concerns by combination of appropriate parts of the preparatory questions. The screen is designed to allow the user to iterate between questions for the different stakeholder concerns to encourage a holistic approach to the evaluation. The final step in Evaluation Planner gives a summary of the user input; presenting an overview of this information allows the user to consider the design at a strategic level, giving them early feedback on whether or not the study will meet their needs. Initial formative evaluation of the toolkit suggests that this summary step does indeed encourage reflective practice (Oliver, McBean, Conole and Harvey, forthcoming).

Evaluation Advisor

This section covers the planning of the implementation aspects of the evaluation. It uses the output from Evaluation Planner as a starting-point, guiding the user through the

Stakeholders:	Lecturer	Lecturer	Student
Concern:	Timesaving	Research	Usability
Question:			
Is this a specific question?	Non-specific ▾	Non-specific ▾	Specific question ▾
Do you want or need to involve all your students in this aspect of your evaluation?	Some students ▾	None ▾	All students ▾
Does this aspect of the evaluation need to be evaluated as part of the teaching process or can it be evaluated in isolation?	In isolation ▾	In isolation ▾	As part of teaching ▾
Save your answers to the questions above		OR	Show All Methods
List of potential data capture methods:	Concept maps, Focus Groups, Unstructured interviews	System log data, Cognitive walk through/think aloud protocols, Designing experiments & pre- and post testing, Resource Questionnaires, Split screen video, Trials	Confidence logs, In-course experiment, Performance test, Reflective logs/student diaries

Figure 4: Examples of the questions used to elicit information from users.

choices of data collection and analysis required in the evaluation process. It links closely to a knowledge base of relevant external material, such as *The LTDI Evaluation Cookbook* and other resources.

Having devised a clear idea of the strategic direction of their evaluation, the Evaluation Advisor section guides the user through the process of identifying appropriate data collection and analysis for the evaluation. After an initial introduction to the section (About the Advisor), the user is asked to provide information about the evaluation questions relevant to the stakeholders identified in the previous section, which they now wish to address (Figure 4).

From this information, appropriate methods for capturing data are suggested by the Toolkit as outlined below (Figure 5).

Name	Definition	Select Method:
System log data	An unobtrusive method for collecting data on student use of software by collecting and recording student activity - such as key strokes, menu selection, and mouse clicks	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO
Cognitive walk through/think aloud protocols		<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO
Concept maps	A visual representation of associated ideas or concepts showing the relationships or links between them.	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO
Confidence logs	Confidence logs involve a series of statements about aspects of a course; students are expected to say how confident they are (using a Likert scale) about each of these. The logs are administered as a survey, and may be used once (in order to get a "snapshot" of class confidence about a particular aspect of the course) or repeatedly (to get a longitudinal profile of how confidence develops over time). Confidence logs were used by the TILT project as part of a suite of methods designed to assess how well learning technology was integrated into teaching and learning. However, the idea of assessing confidence using Likert scales as a part of surveys has been used for many years.	<input type="checkbox"/> YES <input checked="" type="checkbox"/> NO
Designing		

Figure 5: An example of the range of suggestions made by the toolkit.

Behind this process the programme has mapped potential answers to these questions to the different types of data capture method. In this way it is possible to provide the user with a filtered set of suggested methods. However, this filtration does not preclude the users from exploring the wider range of data capture methods. The filtration is designed to provide a layered view of the details about data capture. There is a substantial body of associated information available about each of the data capture methods and this layered approach is designed to help guide the user through this information in a digestible format. For example, some users of the programme will already have a very clear idea of what focus groups can offer as a data capture method, while others will have little background knowledge.

Once confident of the different data capture methods and their uses, users choose which of the methods will be appropriate for their study. Their choice is then used as a starting-point for selecting techniques to analyse any data collected. The next stage again begins with a set of questions designed to explore the purpose of the data collection. In a similar fashion data analytical methods are mapped to the question responses and the users can either choose to accept the returned methods or explore other possibilities in more detail. At the

end of the section there is a summary of the user's input, providing an overview of this information. As previously outlined, this aims to encourage the user to consider the implementation of the evaluation holistically and then reflect on the design.

Evaluation Presenter

This third section looks at the dissemination of the findings of the evaluation to the stakeholders identified in Evaluation Planner. It targets the report format to the audience identified by the evaluation planner and then provides structured templates that take the output from Evaluation Advisor and present it according to the most appropriate media types (Report format, Web site, *PowerPoint* presentation, peer persuasion/oral communication and research-paper format).

Before considering the presentation techniques users work through a 'closing the loop' step. This stage prompts them to build in reflective checks to consider whether their results are valid as well as draw up an evaluation plan. In particular users are asked to think about their evaluation presentation plans in light of the original stakeholders identified in the Planner section and their associated concerns. Relevant information from the previous two sections is then pulled into this section. This stage is important as the toolkit attempts to ensure that users have focused their presentation outputs to reflect the interests of their potential audience. As with the data capture and analysis section, the potential set of presentation tools is prefaced by a set of questions designed to fit the tools to the user requirements. For example, if the requirement is to disseminate the findings of the study quickly, an email or Web site is more likely to be returned as a suggested presentation method than a detailed peer-reviewed journal article. Subsequent questions help the users to draw up a communication plan, considering issues like the message to be communicated, opportunities for dissemination and the resources required.

The final part of Evaluation Planner gives a full summary of all the information given and decisions that the user has taken within all three sections of the toolkit. This acts as an Evaluation Plan for the study. Again this is designed to encourage reflection on the whole process and acts as a valuable one-to-two page summary plan of the process. This can also act as a working document that will allow users to manage the future implementation of the study. The formative evaluation shows that this step is considered a valuable 'reward' by the users after the considerable effort involved in designing the study working through the toolkit.

Further elements of the toolkit

The toolkit also contains background details on the resource itself, the overall architecture, and reasons why users might find the toolkit useful. All the associated resources used throughout the toolkit are grouped in the resource section and can be searched in a number of ways. Users can also add their own resources.

One of the most overwhelming aspects of the feedback from the formative evaluation was that users would welcome a quick route through the toolkit. This is possible via the Evaluation Finder section. Users who have completed an evaluation can choose to make this publicly available for other users to copy and customize. This means that users can pick an off-the-shelf Web site design evaluation and with minimum effort adapt it to their own requirements and local context. It is hoped that a rich database of these evaluation

case studies will be developed. At a more local level, as each user is assigned their own private area of the toolkit they can build up a set of their own evaluation case studies. These can be developed and adapted over time or might be used within a local setting with, for example, a set of colleagues across a department.

Formative evaluation of the toolkit

The evaluation of the toolkit was designed to provide feedback on the usability of the toolkit and to assess its potential impact on practice. A more detailed outline of the evaluation is described elsewhere (Oliver, MacBean, Conole and Harvey, 2000) and only illustrative extracts will be described here. The methodology included observational studies and a follow-up workshop of new users. In the initial stage usability trials in the form of cognitive walk-through were carried out at University of Bristol, University College London and the Dublin Institute of Technology. These were used to improve the design and layout of the toolkit and help the content developers to identify areas that required further work and improvement. Following a think-aloud protocol thirteen participants across the three sites participated in this part of the study. In each case a researcher observed and recorded the user's activities. At the end of the session users also provided feedback on the overall use and value of the toolkit, good and bad aspects, and whether or not the format of the toolkit had influenced the way in which they had approached the evaluation planning process.

Feedback from these trials was used to improve the toolkit and in particular the overall structure and navigation of the resource. Once an updated version of the toolkit had been produced, taking into account all the feedback from the initial usability trials, the second phase of the evaluation was carried out. This took the form of a workshop of 'critical friends' comprising a range of potential users (lecturers, managers, researchers, staff in university support services or national centres). During this one-day workshop, the participants worked through the toolkit section by section. They were asked to keep a record of their activities and in particular to jot down any significant findings or errors encountered. At appropriate points the group were drawn together to discuss progress and in particular good and bad features of the toolkit. The workshop concluded with a general summing up of their experiences with using the toolkit and its potential value and use.

Summary of the evaluation studies

Much early feedback on the toolkit concerned its usability. The navigation, for example, caused problems for some users.

It's not clear to me at what point I go on, so I'm taking the view that there's enough intelligence in here [the toolkit] to stop me if there's something missing. (Workshop)

The 'change selection' buttons at each stage are not clear. (Study U1)

(Note that for this and subsequent quotes, the coding refers to the university where the study took place, U for London, B for Bristol and D for Dublin, and an assigned number for each participant).

Similarly, some users found some of the terminology confusing.

'Evaluation Object' could be replaced by 'Focus of the Evaluation'. (Study D5)

Scope? What do you mean by scope? (Workshop)

It was evident that many of the users were frustrated by the early navigational structure and layout, as this was fundamentally impeding their progress and hampering their understanding of the issues and concepts being described in the program, as the following quote illustrates: 'At the core is probably a sound idea but it is so obscured by the language and interface' (Study D3). The toolkit was iteratively refined in light of such comments, with the result that the usability of the toolkit was markedly improved by the end of the project.

Feedback on the content and potential usefulness of the toolkit was generally positive. Users clearly gained benefit from working through it and recognized that it was a rich resource of material. However, there was some concern that the toolkit was deceptive in terms of its size and ease of use. In particular, it was recommended that users should be made more clearly aware of the time required to complete each section (approximately one hour) and the level of detail and concentration required if the user was to gain optimal value from using the resource.

I really liked [it] making me think about the purposes of evaluation. I've completely changed my view of the evaluation by working through this. (Study B3)

I haven't had to produce an evaluation plan ever before – so in that sense it was extremely helpful as it guided me through the process, explained some background ideas and suggested other sources of help. (Study U2)

Further feedback was derived from the 'critical friends' workshop and the final steering group meeting. Some users expressed concern that, to be utilized effectively, the user really needed to spend a good few hours working through the toolkit and that for smaller evaluations this was impractical. In response to this, simpler 'quick' routes through the toolkit, along with templates for standard evaluation processes, are planned. However, overall, feedback has been positive and the general consensus was that the toolkit had the potential to be a very valuable resource. The follow-on workshop echoed many of the comments made in the observational studies. It was encouraging to note that the improvements made to the toolkit, as the result of initial feedback, were noticeable in the way that the workshop users worked through the resource much more easily, although further improvements to the navigation structure will still need to be made. Perhaps most encouraging was a general view that the toolkit encouraged reflective practice and that it would build into a valuable resource for individuals as they developed a library of their own evaluation plans for a variety of activities. The workshop organizers also noted that the use of the toolkit as a framework for an evaluation workshop significantly increased the degree of evaluation knowledge imparted to participants. (In related face-to-face workshops based on a paper version of a precursor to the toolkit, participants generally only managed to work through a third of the material in the same amount of time.) Equally important was the feedback indicating that participants valued the experience of going through the various development stages involved in creating their own plans with the option of being able to reuse and adapt plans developed by other users of the toolkit. The ability to store their thought processes and answers to different sections online for consideration at a later date was also particularly welcomed.

Conclusions

The work described in this paper has outlined the development of an online evaluation toolkit, which is designed to provide easy-to-use guidance and help for practitioners interested in developing evaluation plans to support aspects of their learning and teaching. It has described the philosophy behind the development of the toolkit, the component sections of a toolkit for evaluation, and feedback from a formative evaluation of this resource.

Feedback from the trials was positive, with many of the users reporting that the toolkit helped guide them through their evaluation process as well as provide them with useful links to further information and support. However, these initial studies demonstrated the need for additional trials and development work, as well as a series of workshops to introduce the resource to potential groups of users. Other areas of work suggested could include:

- case studies covering 'standard' types of evaluation, perhaps provided by subject experts acting as consultants. The consultants would attend an initial workshop on using the toolkit and then use the toolkit to generate the case studies. Studies of interest include the evaluation of Web sites, externally funded projects, a learning and teaching innovation, a strategy document, a resource or a Virtual Learning Environment;
- user trials to build on the limited but valuable information gained during the short time-scale of the project. Studies with a more diverse group of users, concentrating on the ways that they use the resource and the aspects they find most useful, would improve the value and relevance of the toolkit itself. They would also help define the key factors for success in producing toolkits and hence help define specifications for future related resources of this kind;
- adaptation of the existing resource for other user groups, such as further education;
- extension of the toolkit to cover other areas of learning and teaching such as curriculum development, media selection, assessment or quality assurance.

Toolkits clearly represent a valuable type of resource for staff such as educational developers and learning technologists who work and support academics engaged in the development of learning and teaching practice. The work described within this paper would support the idea that the development of an online toolkit is an effective way of supporting the planning phases involved in carrying out an evaluation study.

References

- Conole, G. and Oliver, M. (1998), 'A pedagogical framework for embedding C and IT into the curriculum', *ALT-J*, 6 (2), 4–16.
- Conole, G., Oliver, M. and Harvey, J. (2000), 'Toolkits as an approach to evaluating and using learning materials', *ASCILITE 2000*, Coff's Harbour, NSW, Australia.
- Harvey, J. (1998), *The LTDI Evaluation Cookbook*, Learning Technology Dissemination Initiative.

HEFCE (1996), 'Evaluation of the teaching and learning technology programme', Coopers and Lybrand report, HEFCE M21/96, http://www.niss.ac.uk/education/hefcel/pub96/m21_96.html.

HEFCE (1997a), 'Teaching and Learning Technology Programme Phase 3: invitation to bid', HEFCE, 14/97, TLTP, Bristol.

HEFCE (1997b), 'Information technology assisted teaching and learning in higher education' (ITATL report), Bristol: HEFCE Research series.

Oliver, M. and Conole, G. (1998), 'Evaluating Communication and Information Technologies: a toolkit for practitioners', *Active Learning*, 8, 3–8.

Oliver, M. and Conole, G. (1999a), 'From theory to practice: a model and project structure for toolkit development', BP ELT Report No. 12, University of North London.

Oliver, M. and Conole, G. (2000), 'Assessing and enhancing quality using toolkits', *Journal of Quality Assurance in Education*, 8 (1), 32–7.

Oliver, M., MacBean, J., Conole, G. and Harvey, J. (forthcoming), 'Using a toolkit to support the evaluation of learning technology', *Journal of Computer Assisted Learning*.

Phelps, J., Oliver, M., Bailey, P. and Jenkins, A. (1999), 'The development of a generic framework for accrediting professional development in C&IT', EFFECTS report No. 2, University of North London.