Information and communication technologies and their role in flexible learning: AUTC project

Review of evaluation frameworks

Ron Oliver, Catherine McLoughlin, Jan Herrington

Please note: This report was conducted as a research task in this project to assist in the development of the project's evaluation instrument: Evaluation and Redevelopment Framework. It is to be considered as a work-in-progress.

Introduction

As a result of Workshop 1, the task was set to 'search for existing relevant instruments that may inform and/or be incorporated into the Evaluation and Redevelopment Framework (ERF)'. This paper provides a summary of the findings that emerged from an exploration of existing instruments, frameworks and checklists, and a description of the parameters of the task. The exploration sought to investigate evaluation instruments of technology-based learning environments, although some non-technology frameworks were also considered where they had readily been adapted to multimedia or online learning environments.

The approach used adopted three broad strategies:

1. Experts were contacted direct and asked to nominate 'any instruments that currently exist that have been used to evaluate technology-based learning products, particularly instruments that explore the learning quality or learning potential (as distinct from interface etc)'. Experts contacted included: Curtis Bonk, Betty Collis, Jim Taylor, Tom Reeves and Gerry White.

2. Journals and known published frameworks, including a special edition of *Higher Education Research and Development* (Vol 18, No. 2, 1999, Edited by John Bain) devoted to evaluations of innovations in higher education.

3. An extensive search of internet resources using search descriptors such as: evaluation, instrument, checklist, questionnaire, online learning, technology-based learning, learning activities, evaluation toolkits etc.

The frameworks were reviewed, and summarised across the following criteria:

Name of framework and creator

Purpose of framework

Methodology & data sources

What aspects of student learning does the framework consider?

What aspects of technology delivery does the framework consider?

Relevance to ERF/limitations

The descriptions of a selection of frameworks are provided as appendices to this summary document.

The process of reviewing the frameworks was not conducted in a 'grounded' manner which sought to assemble a comprehensive list of elements which might comprise an ideal evaluation framework. Rather, each frameworks was evaluated on two counts: firstly, whether it could usefully act as a tested and validated substitute for the ERF; and secondly, whether any important elements of the framework had been neglected or omitted in the proposed ERF (revised version).

Summary of frameworks

The purposes of evaluation frameworks are varied, many designed to serve a specific and limited function. It is possible to group the frameworks that were evaluated for this exercise into several categories for the purpose of the summary. For example, one method is to group the frameworks into four broad types:

1.   Evaluation of educational innovation projects

Some frameworks focussed on the evaluation of the process of the completion and implementation of an educational innovation or project. Such frameworks cover the whole life of a project, from its origin, design, context, development, resource selection, implementation, evaluation and outcomes (e.g., Alexander, 1999, Gunn, 1999)

2.   Evaluation of websites (Checklists)

There were literally hundreds of sites, predominately on the Internet, with frameworks for the evaluation of instructional, and other, websites. The majority of these were in a checklist format. Such frameworks considered aspects of websites such as the accuracy and currency of information, the authority of the site, the use of multimedia, ease of navigation and access, and the appearance of the site (e.g., Beck. 1997).

Generally, the value of checklists has been called into doubt for a number of reasons:(See Tergan, 1998). These include:

A lack of reliability (Correlation of ratings between reviewers is usually low.)

Concerns about validity (Criteria are often based solely on the consensus of developers.)

Bias.(Themes important to one group of stakeholders, but not otthers)

The aggregation of scores (The weighting of categories of questions is typically neglected in scored systems.)

Neglect of individual differences. (Judgement is based on the reviewer's opinion, rather than on a cross-section of students' opinions.)

Neglect of context (Many reviews focus on the software as an object, and not on its use in a particular context.)

The lack of tailored criteria (The categories developed in one context may not be relevant in another.)

3. Effectiveness of learning frameworks

Some frameworks focussed strongly on learning outcomes, and entailed questions and checklist on whether learning environments promoted higher order thinking, collaboration, reflection, motivational aspects of learning, authenticity, creativity and so on (e.g., Reeves and Laffey, 1999). Some of these were originally developed for specific learning situations but had been adapted to other technology-related settings (e.g., Chickering and Ehrmann, 1994). It is worth noting that there were few concrete suggestions of how to evaluate these aspects of learning beyond prompting evaluators to consider them, for example, 'process engages the learner', 'challenges learners to think' etc. However, most of these authors have advocated triangulation of findings in order to increases the credibility of findings. Some were more specific, such as 'unit requires synthesis of multiple sources of information, and/or taking a position, and/or going beyond the data used and making a generalization or creative product' (e.g., Center for Research on Learning and Technology, 2000).

4. Evaluation Tools for practitioners (Toolkits and Handbooks)

The need for practitioners to carry out their own evaluations has led to concerns about expertise in conducting evaluation. Lecturers, for example, may have expertise in their discipline and in teaching, but it is unreasonable to assume that they will have expertise, training, and in many cases even experience of carrying out program evaluations (Oliver & Conole, 1998a). As a consequence, several tools have been developed to support practitioners engaging with evaluation. *The Evaluation Cookbook* (Harvey, 1998), for example, provides a series of evaluation 'recipes', each summarising a methodology in an easy to follow form, complete with hints and tips. Whilst the cookbook provides a 'how to' guide for implementing evaluation studies, *The ELT Toolkit* (Oliver, 1999d) focuses on their design. It is structured around a model of evaluation design that incorporates six stages, with the first two and the last relating to the context, and the middle three focusing on the details of the study itself. These steps are as follows (Oliver et al., 1998):

Identification of stakeholders

Selection and refinement of evaluation question(s), based on the stakeholder analysis

Selection of an evaluation methodology

Selection of data capture techniques

Selection of data analysis techniques

Choice of presentation format

At each step, users complete activities that support their use of a structured knowledge base (consisting of descriptions of methods). It is argued that this helps them make informed decisions that would otherwise be beyond their existing level of expertise. The toolkit is based on the pragmatic framework for methodology use outlined above, in that no single approach is presented as being correct. Instead, all are described in terms of their distinguishing characteristics, and the purpose of the exercise is to allow practitioners to select the approach best suited to their current situation.

An integrated approach that covers both design and application is the Flashlight program. *The Flashlight Project* (Erhmann, 1999) has produced a questionnaire-based toolkit that provides a simple structure for evaluation by practitioners. The tool is based on an analysis of three elements:

A technology

An activity for which it is used

The educational outcome of the activity

The bulk of the tool is concerned with identifying questions that can be used to generate data. This includes consideration of five thematic prompts (e.g. "questions about the use of the technology"), and access to the Current Student Inventory, a repository of questions devised by other users. The Flashlight tool is based on the premise that "very different educators need to ask similar questions" (Ehrmann, 1999). This position has been challenged as an over-simplification (Oliver & Conole, 1999). This assumption allows the Flashlight project to justify a focus on only one type of methodology, but it ignores the variety of more appropriate methodologies that the ELT toolkit, for example, seeks to identify (Oliver & Conole, 1998). A comparable approach involves the creation of handbooks, such as that of the CUTSD project (Phillips et al., 2000). This resource also incorporates design and implementation guidelines, and focuses on one particular approach (action inquiry). However, unlike the Flashlight tool, no claims are made to general applicability. The handbook contains information and advice that would be of use to any practitioner engaged in action research.

5. Other

A few other frameworks which do not fit into these broad categories include those that have been developed for a specific context, such as: peer review of quality of ICT-based teaching and learning services (Taylor, 2000); evaluation of course management software, such as *Blackboard* (Institute for Higher Education Policy, 2000) and the **SoURCE Evaluation Framework** (Twining, Blake & Taylor, 2000). This approach is to identify existing educational software that has been shown to be effective in one context, and evaluate the process of customising it for reuse in another context. The main focus of SoURCE is to take software that has already been successfully designed , implemented and evaluated in one context and customise it for others. Customisation means that any reuse of software includes changes to the software and/or pedagogy. As software is seen to be embedded in one context, it is not not assumed that customisation will avoid changes to both software and pedagogy. This is relevant to the ERF Phase 3.

Some elements of evaluation are not explicitly stated in the current version of the ERF, for example: motivation for learning, scaffolding, and expert performance. However, these may be present implicitly, or could be further developed in the sub-points of the ERF.

References

Beck, S. (1997). *Evaluation criteria: The good, The bad & the ugly: or, why it's a good idea to evaluate web sources.* Institute for Technology-Assisted Learning, New Mexico State University. Available: http://lib.nmsu.edu/instruction/eval.html

Center for Research on Learning and Technology (2000). *Rubric for evaluating TICKET (Teacher Institute for Curriculum Knowledge about Integration of Technology) curriculum integration projects*, Indiana University. Available: http://www.indiana.edu/~tickit/index.htm

Chickering, A.W., & Ehrmann, S.C. (1994?). *Implementing the seven principles: Technology as lever.* American Association for Higher Education. Available: http://www.aahe.org/technology/ehrmann.htm

Gunn, C. (1999). They love it, but do they learn from it? Evaluating the educational impact of innovations. *Higher Education Research and Development, 18*(2), 185-199.

Institute for Higher Education Policy (2000). *Quality on the line.* Study commissioned by the National Education Association and Blackboard Inc. Available: http://www.ihep.com/PR17.html

Senate ad hoc Committee on Teaching Quality, Effectiveness and Evaluation (1998/99). *Effective teaching principles and practices.* University of British Columbia. Available: http://wwwcstudies.ubc.ca/facdev/services/appendixc.html.

Name of framework and creator	Purpose of framework	What information does it elicit?	What aspects of student learning does it consider?	What aspects of technology delivery does it consider?	Relevance to ERF/Limitations
Situated evaluation of computer-assisted learning Gunn	Evaluation of the effects of technology, not in isolation, but within the context of the whole learning environment (including integration, institutional and cultural factors)	Recommends a variety of data sources to evaluate innovation such as observation, independent review, expert review, questionnaire, assessment results, discussion and log data	Learning objectives and measurement of achievement Motivational factors (attention, relevance, confidence and satisfaction)	Hardware and software issues (functionality, standards, intellectual property) Implementation of technology (is it effective?)	Motivational factors addressed?
Bain, Alexander & Hedberg as used in CUTSD 1999 at : http://cleo.murdoch.edu.au/projects/cutsd99/	Holistic framework that is learner centred, covering both formative and summative evaluation of the context and learning process	Same as Gunn ie data triangulation, at both formative and summative stages	Impact of the innovation on learning process, student thinking processes, confidence, understanding, interpretation and reflection	Attractiveness, accessibility useability and functionality	Is learning centred and holistic, and also contextualises the evaluation within curriculum .Is consistent with B & P guidelines. Additional dimension or 'enhancing attribute' could be curriculum analyses & student perceptions of learning context
Reeves & Laffey, 1999 (HERD, 1999, vol18, 2, 219-233)	Set of 14 pedagogical dimensions to guide the design and evaluation of effective interactive learning systems	Dimensions defined as by-polar continuum of values. The extremes are not readily applicable but serve as indicators	Goal orientation, pedagogy, instructor role, role of technology, nature of learning activities, source of motivation, experiential validity, cooperative learning,	Only technology as instructivist or constructivism ie technology use as 'surrogate instructor' or 'cognitive tool'	Compatible with B& F framework. Is 'source of motivation' an additional attribute?

Name of framework and creator	Purpose of framework	What information does it elicit?	What aspects of student learning does it consider?	What aspects of technology delivery does it consider?	Relevance to ERF/Limitations
Erhmann (1999) Flashlight Evaluation Handbook Report at http://www.learner.org/edtech/rscheval/flashlight/toc.html	Applies the seven principles and in addition provides a Flashlight Evaluation handbook containing student inventories	Tool kits that consist of: Student inventory Faculty inventory Cost analysis manual Learning outcomes inventory Superior Inventory Distance education staff inventory Authoring and remote data analysis systems	Active learning Collaboration High expectations Rich & rapid feedback Engagement cognitive & creative outcomes Respect for diversity Application to real world	Access, learning outcomes of technology use, cost	Compatible with B&B P framework but is more outcomes focussed, ie seeks evidence of **learning outcomes**
Taylor (2000)	Peer review of quality of ICT based teaching & learning services	Peer review Quality standards Evidence scholarship of teaching	Not applicable	Effectiveness Sustainability	The conceptual framework is different

Name of framework and creator	Purpose of framework	What information does it elicit?	What aspects of student learning does it consider?	What aspects of technology delivery does it consider?	Relevance to ERF/Limitations

Rubric for evaluating TICKET (Teacher Institute for Curriculum Knowledge about Integration of Technology) curriculum integration project
Indiana University (Curtis Bonk)
	Evaluation tool for projects incorporating technology in the classroom	The tool is a rubric giving points (3, 4 or 5 points) for how well the project satisfies a range of criteria.	Authenticity and intellectual challenge :
			  Fosters higher order thinking skills
			  Takes students beyond their classroom
			  Significance, intellectual challenge and motivating quality of the content
			Evidence of student learning:
			  Convincing examples of student work
			Creativity
			  Creativity and originality of project ideas and materials
				Technology infusion:
				  Extent to which appropriate technology has been incorporated into the curriculum unit
				Design quality of student materials:
				  Extent and quality of resources and references provided to students
					Attempts to define concrete elements of evaluation of learning

Effective teaching principles and practices University of British Columbia	A set of principles and practices against which to evaluate teaching activities	The tool is a checklist of principles to be used as a model for departments and faculties, to reflect the highest standards of classroom teaching	Principle 1: Sets clear goals and intellectual challenges for student learning Principle 2: Employs appropriate teaching methods and strategies that actively involve learners Principle 4: Attends to intellectual growth of students Principle 6: Incorporates learning beyond the classroom	No specific references to technology. General teaching principles.	General principles compatible with ERF

Name of framework and creator	Purpose of framework	What information does it elicit?	What aspects of student learning does it consider?	What aspects of technology delivery does it consider?	Relevance to ERF/Limitations
Seven principles of good practice (Technology as lever) Chickering & Ehrmann	Evaluation of the use of technology to enhance good practice in undergraduate education	Framework is a checklist against which to assess good practice. The authors recommend observation to assess the impact of the approach on learning outcomes.	Reflective communication Collaborative learning Application of learning in daily life Multiple perspectives of learning Enhancing diverse talents and ways of learning	Communication technologies (email, internet discussion boards, computer conferencing etc) Computer simulations The use of technology (such as video, computers, internet) to: track progress, provide feedback, provide storage and easy access, increase efficiency and provide a variety of perspectives Use of technology (internet) to limit constraints of time and place	General principles compatible with ERF

Name of framework and creator	Purpose of framework	What information does it elicit?	What aspects of student learning does it consider?	What aspects of technology delivery does it consider?	Relevance to ERF/Limitations
Quality on the line National Education Association (NEA) and Blackboard Inc.	List of quality benchmarks for the evaluation of online distance learning units in higher education	The framework comprises a list of 24 quality measures (based on practical strategies from US colleges considered to be leaders in online distance education). The benchmarks distilled from this study were divided into categories of quality measures.	6. Courses are designed to require students to engage themselves in analysis, synthesis, and evaluation as part of their course and program requirements	All items apply to online learning environments: 1. A documented technology plan that includes electronic security measures to ensure both quality standards and the integrity and validity of information. 2. The reliability of the technology delivery system is as failsafe as possible. 14. Students receive information about programs, including admission requirements 18. Technical assistance in course development is available to faculty, who are encouraged to use it.	Not particularly relevant. Used to evaluate course management software rather than learning designs
Web evaluation criteria Institute for Technology-Assisted Learning, New Mexico State University	Evaluation tool for the quality of educational websites	Checklist of 17 items against which to evaluate the quality of an online learning unit There are hundreds of evaluation sites like this one for web-based units. This one groups checklist items under headings: Accuracy, Authority, Objectivity, Currency and Coverage	No learning outcomes are addressed	Checklist items apply solely to websites.	Not particularly relevant, evaluate general presentation of websites rather than learning designs or activities