Evaluation of the TELT platform

Essential elements and methodologies

[DRAFT FOR DISCUSSION]

Stephen Quinton  
Mariya Pachman  
Richard He

L&T@UNSW  
January 25, 2010

This document is a collection of reports explaining the administrative, technological and methodological factors to be observed in the development and implementation of a new evaluative framework for technology-enhanced teaching and learning (TELT) at UNSW.
# Table of Contents:

**Summary of Steps for Development of the TELT Evaluation Framework**........................................3

**Part I. An Iterative Model for the Evaluation of TELT Platform Applications (summary)**............................7

   References.............................................................................................................................................13
   Diagrams...............................................................................................................................................15

**Part II. Evaluation of Technology-enabled Teaching and Learning**.........................................................19

   Advantages and drawbacks of the existing evaluative frameworks..............................................20
   Biases..................................................................................................................................................23
   Open Source evaluation....................................................................................................................28
   Attributes of a Generic Evaluation Framework..................................................................................29
   Integrative models and their shortcomingss ......................................................................................31
   Summary Conclusions.......................................................................................................................32
   Tables...................................................................................................................................................33
   References.............................................................................................................................................35

**Appendix 1: Review of Methodologies for evaluating technology-enabled learning and teaching**........39

   The Pedagogy Layer.........................................................................................................................40
   The Technical Layer ........................................................................................................................89
   The Business Layer ........................................................................................................................90
   Conclusion..........................................................................................................................................92
   References..........................................................................................................................................93

**List of tables and figures:**

   - Table 1...............................................................................................................................................42
   - Table 2...............................................................................................................................................48
   - Table 3...............................................................................................................................................52
   - Table 4...............................................................................................................................................57
   - Table 5...............................................................................................................................................61
   - Table 6...............................................................................................................................................67
   - Table 7...............................................................................................................................................72
   - Table 8...............................................................................................................................................73
   - Table 9...............................................................................................................................................78
   - Table 10............................................................................................................................................78
   - Table 11............................................................................................................................................80
   - Table 12............................................................................................................................................86
   - Table 13............................................................................................................................................88
   - Figure 1............................................................................................................................................41
   - Figure 2............................................................................................................................................75
   - Figure 3............................................................................................................................................77
   - Figure 4............................................................................................................................................84
   - Figure 5............................................................................................................................................85
   - Figure 6............................................................................................................................................88
Summary of Steps for Development of the TELT Evaluation Framework

1. Establish rationale and argument for devising a generic evaluation framework:
   - lack of availability of a widely recognised framework
   - the need to avoid the many bias inherent in most frameworks
   - the need to position L&T at the top of a hierarchy of evaluation criteria consistent with the TELT vision
   - identify the principles and models on which the framework is to be established.

2. The evaluation framework to consist of three stages:
   - Vendor – Stage 1
   - Pre-pilot scoresheet – Stage 2
   - Full pilot evaluation – Stage 3.

3. Stage 3 of the Framework incorporates three layers of evaluation criteria:
   - L&T (incorporating the student voice)
   - Business processes
   - Technical Requirements.
   The preceding order to be established to emphasise the fundamental importance of L&T in determining the relevance of business processes technical requirements for serving the vision and needs of the TELT Platform.

4. Three subcategories (sub-layers) of L&T evaluation apply to the L&T layer:
   1. the educational value and usefulness of a technology application for inclusion in the TELT platform
   2. the learning effectiveness of approved TELT applications, and,
   3. the learning environments that are supported by approved TELT applications.

5. Sub-layer 1 observes several key sources of evaluation priorities:
   - L&T criteria that are informed by and contribute to the University’s Graduate Attributes, the UNSW L&T principles and guidelines, and the 16 UNSW Guidelines for L&T practice
   - L&T criteria specific to faculties and schools preferred teaching models / learning approaches applied as a basis on which to develop the L&T evaluation criteria, and
   - L&T criteria tailored to the type and scale of the technology (category) to be assessed.

6. Categories of TELT Applications:
   To permit relevance to different types of TELT applications (categorised for example as large-scale LMS, medium-scale virtual classroom or collaborative suites, and small-scale special purpose software), evaluation methodologies are to be identified to address the specific requirements of each category.
7. Evaluation methodologies are currently being identified for:

- All three stages of the evaluation framework (Diagram 1)
- All three sub-layers of Stage 3 L&T evaluation criteria (Diagram 3), and
- All methodologies to be refined /adapted to match the type / scale of TELT application under evaluation.

Steps 1 to 7 to lay down a foundation for:

- Research on the effectiveness of the TELT evaluation framework
- Research on the learning effectiveness and the usability of TELT applications
- Research on the impact of the TELT Platform on the educational effectiveness of learning environment design
- Publication of Journal papers / book chapters
- Preparation of survey instruments for L&T (staff, students), business processes, and technical requirements
- Reports (covering both formats and audiences)
- Informing TELT training plans and processes
- Informing digital resource design and developments
- TELT policy, plans, guidelines (for both TELT Platform and TELT portfolio levels).

To be determined at the appropriate time:

- Data storage
- Data coordination
- Data access / sharing
- Reporting formats, types, levels and audiences
- Wizard for selection of evaluation criteria.
Decision Analysis Map to plan the TELT Evaluation Framework

Field of TELT evaluation is still not saturated: new models appear

- Inherent biases (in particular, technology-driven evaluation)
- Need for an iterative component
- Multiple values models
- Existing models are prescriptive
- Narrow evaluative scope + network effects of technology
- Theory base

why?

Informatization of society
Evaluators background
Pluses and minuses of quality assurance benchmarks

why?

TELT evaluation requires new evaluative approaches

needs

Open Source promotes democratic knowledge construction process

Open Source evaluation

Future of educational technologies

Educational technologies as a stand-alone pedagogical element

Learner-centeredness

Avoiding an additive approach (1+1 technology) to accommodate the learner and to support the organizational change

If admin component is well-defined it will help finding holes in organizational structure

Philosophical part

Pedagogical, technical and organizational aspects

Possible difficulties with methodology

"Worldware" (e-mail, Internets, Office)

Learning and technology objectives, also institutional strategy and policy

include

Learner as collaborator

Educational technologies are a campus-wide systems integrated with enrolment, library services etc.
## Development Timelines (suggested)

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/1/09/09</td>
<td></td>
</tr>
<tr>
<td>1/12/09</td>
<td></td>
</tr>
<tr>
<td>1/02/10</td>
<td></td>
</tr>
<tr>
<td>1/04/10</td>
<td></td>
</tr>
<tr>
<td>1/06/10</td>
<td></td>
</tr>
<tr>
<td>2/1/09/10</td>
<td></td>
</tr>
</tbody>
</table>

**Phase 1: develop argument for a framework design**

- Phase 2: framework design (micro and macro scales)
  - Aims, objectives, outcomes
- Phase 3: adapting an evaluation methodology
  - Developing an evaluation plan
- Phase 4: framework refinement (existing teaching practices and approaches)
  - Developing evaluation rubrics
- Phase 5: commence evaluating the TELT platform

Ongoing Recurrent stage: reporting & presentations (feeds back into all preceding stages)
An Iterative Model for the Evaluation of TELT Platform Applications

Summary

An extensive literature search conducted over several months has confirmed the difficulty in identifying a comprehensive, generic approach and methodologies for facilitating the selection and evaluation of the effectiveness of educational technology applications and solutions that focus exclusively on serving the needs of learning and teaching (L&T). In the UNSW context, a well-designed, highly flexible evaluation framework is crucial to ensuring all software applications and tools selected for the TELT platform will meet the demands of L&T both now and in the future.

In developing such a model, the overarching vision is to devise a ‘living’ evaluation framework that affords an iterative refinement process to underpin: all TELT technology selections and developments; efficient tracking of the trajectory of success (and failure) of educational technologies used across faculties and over time; and, enhancement of the divergent L&T approaches and practices in online learning at UNSW.

Within the proposed model, the evaluation framework in itself does not escape attention. An integral component of the framework is to assess how well the evaluation processes incorporated within each stage assist not only to identify suitable technology applications and tools for inclusion in the platform, but also improve technology supported learning and teaching.

At its core, the evaluation framework comprises three stages. Stage 1 is specific to an assessment of the stability and viability of prospective product vendors. Stage 2 incorporates a pre-pilot scoresheet to determine the suitability of an application for full pilot testing in Stage 3. Stage 3 begins an ongoing L&T focussed evaluative sub-cycle that aims to facilitate a culture of evidence-based excellence at the faculty and institutional levels to improve UNSW technology supported L&T practices.

Foundations of the TELT Evaluation Framework

A three month review of the available research literature has failed to uncover the existence of a comprehensive, generic approach to the evaluation of educational technologies that is guided primarily from a L&T standpoint. For UNSW, the priority for its Technology Enhanced Learning and Teaching (TELT) evaluation strategy is to devise a theoretical framework from which to assess the suitability of selected educational technologies for inclusion in the TELT platform. Once a technology has passed through a three stage evaluation process and approved for general use, a perpetual, iterative process of evaluation is initiated to assess the educational effectiveness of that technology in terms of the depth of student comprehension and the measureable improvements to the teaching and learning processes.

[Refer to Diagram 1]

In essence, it is proposed that the TELT Evaluation framework incorporates three distinct, yet fully interrelated stages. Stage 1 focuses on vendor assessment; Stage 2 involves a pre-
pilot scoresheet test; and Stage 3 initiates an ongoing evaluative cycle, first to ensure the targeted technology application (or solution/s) complies with the established criteria (and subsequent inclusion or exclusion from the TELT platform) and second, to evaluate the learning effectiveness of the approved educational technology to align with the need to enhance and improve online L&T practices at UNSW. The Stage 3 cycle also feeds into a broad-level cycle of evaluation of the entire TELT Evaluation framework to ensure continual refinement and modification of all three stages in terms of their respective criteria, their specific interrelationships, and their interdependencies with learning and teaching.

In **Stage 1**, prospective vendors or open source partners are requested to produce evidence of their viability, business processes and long-term commitment to the product. Vendors are assessed using a multi-stepped screening approach. Subsequent potential implementation in the TELT platform is conditional upon successful completion of the pre-pilot (Stage 2) and full pilot (Stage 3) evaluations. The rationale behind the Stage 1 evaluation is to optimise the university’s expenditure and time commitments in determining the suitability of a vendor’s product prior to the Stage 2 and Stage 3 evaluations.

**Stage 2** involves a ‘pre-pilot’ evaluation of a vendor’s or an open source software solution. The process is informed by the outcome of Stage 1 in terms of determining the types and comprehensiveness of the criteria to be applied. A pre-pilot scoresheet (quantitative output) has been designed to assess and compare products (if more than one shortlisted). All products are evaluated against existing TELT Platform needs and planned directions with respect to the range of features offered (functional), the implications for effective integration with UNSW’s business processes, and the product’s technical specifications for detailed comparison with the university’s IT infrastructure requirements. If the product (or products) scores highly, it is selected for the full pilot evaluation in Stage 3.

The rationale for a pre-pilot evaluation is the rapid rate of technological change. The situation when “once something is ...evaluated... it is already obsolete” (Bruce, 1999) began to influence the educational technology evaluation field from the outset of the 1990s. Moreover, the results of longitudinal studies in this area often became redundant by the time a technology solution was ready for implementation, again due to the rapidity of technological advancements and innovation (Baker and Herman, 2003). Stage 2 is designed to mitigate these problems.

**Stage 3** determines the suitability of online learning products or software solutions for inclusion in the TELT platform and subsequent release to the university community. It introduces a three-layered, full pilot evaluation that requires several evaluative aspects to be incorporated into a unified process. On one level, it observes a systematic approach to a detailed evaluation of three crucial elements: the pedagogical, business processes, and technical requirements of the TELT platform. On a much broader level, the framework affords genuine systemic flexibility in selecting the types and complexity of criteria, and the analysis methodologies / approaches that are applied to each of the preceding three elements. The evaluation processes outlined here parallels the approach embraced by Marshall (2007) in the construction of his eLearning Maturity Model (eMM). In describing the "Dimensions of capability" of the eMM model, Marshall (p 6) writes:
A key development that arose from the evaluation of the first version of the eMM is that the concept of levels used was unhelpful (Marshall and Mitchell, 2006). The use of levels implies a hierarchical model where capability is assessed and built in a layered way. The key idea underlying the dimension concept in contrast, is holistic capability. Rather than the model measuring progressive levels, it describes the capability of a process from synergistic perspectives. An organization that has developed capability on all dimensions for all processes will be more capable than one that has not. Capability at the higher dimensions that is not supported by capability at the lower dimensions will not deliver the desired outcomes; capability at the lower dimensions that is not supported by capability in the higher dimensions will be ad-hoc, unsustainable and unresponsive to changing organizational and learner needs.

As indicated, the evaluation criteria applied in Stage 3 consist of three interrelated layers: Pedagogy (L&T), Business, and Technical. The layers (illustrated in Diagram 2) collectively serve to establish a multi-referenced foundation on which to systematically evaluate targeted technology applications and solutions. Built into the first layer of Stage 3 (Pedagogy Layer) is an iterative (three) sub-layered evaluation process to ensure continual modification and refinement of all criteria used in each of the three layers.

[Refer to Diagram 2]

The rational for adopting a three-layered approach is that the pedagogical, technical and organisational aspects represent the fundamental sets of factors required to ascertain the suitability of any learning technology (Console, 2004). As the evaluation framework evolves through further iterations of refinement and modification, it is likely that additional layers may be required. For the present, the three layers recommended here provide an absolute minimal, albeit substantial basis on which to establish the TELT evaluation framework.

The Pedagogy Layer is emphasised first and foremost as it directly supports the visions and principles drawn from the UNSW Graduate Attributes, Learning and Teaching Principles and the 16 Guidelines for Learning and Teaching Practice. To ensure full representation of all UNSW L&T perspectives and needs, the pedagogical evaluation criteria used in this layer is further informed by the preferred teaching approaches and models applied by faculties and schools.

The Business layer reflects the formalised processes of organisational change at UNSW that are informed first by pedagogical and then by technological factors. Any new propositions related to institutional business needs and procedural changes that emerge during the evaluation cycle inform the refinement of the business and technical layer criteria as determined appropriate.

The Technical Layer reflects the need to account for the technical requirements, standards and protocols established at UNSW that are informed first by the pedagogical and then the business process criteria. Any new propositions related to technological advancements and organisational changes that emerge during the evaluation cycle inform the refinement of the technical and business layer criteria, again as determined appropriate.

A common problem encountered in educational technology evaluation is the lack of models that account for both the learning processes and the influences exerted by technology on
the learning process (Attwell, 2006). The European experience demonstrates that the managerial (or business) component is usually well addressed in their models, but lack a primary focus on pedagogical factors. In essence, there is a cogent need to encapsulate all aspects of educational evaluation within generic framework.

[Refer to Diagram 3]

As illustrated in Diagram 3, the Stage 3 L&T criteria (Pedagogy Layer) comprises three iteratively evaluated sub-layers. Sub-layer 1 provides the basis for determining the criteria for the pedagogy layer, which in turn inform the evaluation criteria used by the business and technical layers. Ongoing refinement of the pedagogy layer criteria is achieved through continual reference to sub-layers 2 and 3. In general, sub-layer 1 observes several key sources of evaluation priorities:

a) **L&T criteria that are informed by and contribute to the University’s Graduate Attributes, the UNSW L&T principles and guidelines, and the 16 UNSW Guidelines.** The rationale for including L&T criteria derived from University level vision statements and guidelines is that it may 1) inform the organisation’s L&T structure, procedures, and strategies, and thereby reveal hidden or assumed gaps, and thereby identify the differences between the planned and the existing structures as emphasised in the examples provided by Ehrmann (1995), and 2) assist to merge the crucial pedagogical component into an explicit model for effecting organisational change within tertiary education (Boys, 2002).

b) **L&T criteria specific to faculties and schools.** The rational for including L&T criteria specific to faculties and schools is that any contemporary evaluative model should address the divergent needs of staff and students (Stufflebeam and Shinkfield, 2007), and embed multiple values in the evaluation process to avoid the “whose values” dilemma (Zammuto, 1982).

c) **Tailored L&T criteria appropriate to the type and scale of the technology to be assessed.** The rational for acknowledging the need for tailored selections of L&T criteria relevant to the type and scale of the technology [enterprise wide (LMS, CMS, Moodle, Sakai); collaboration (Connect, Wimba); publishing websites (blogs, wikis); curriculum tools (mapping, management); course content development tools; school or discipline specific applications (Maple)], is that the notion of scholarship varies from discipline to discipline. For example, the original and unique vision of the world and the (general) solitary nature of the work conducted in the arts and humanities are in sharp contrast with the natural sciences research landscape where team work and research group settings are an accepted part of their normal routine. Thus, it must be acknowledged that all faculties have different needs as to what specific applications / technologies most benefit their case.

Sub-layers 2 and 3 do not occur until after the sub-layer 1 evaluation process is completed and the application is subsequently approved for release through the TELT platform. It suffices to note however, that sub-layers 2 and 3 provide the catalyst for an iterative evaluation cycle for all three Stage 3 L&T sub-layer criteria. Given the differing educational emphases and outcomes expected from each sub-layer, the applied evaluation criteria and methodologies must also be tailored to suit.
A closer examination of the Pedagogy layer in Diagram 3 illustrates the three L&T evaluation sub-layers. Once a technology has been approved for inclusion in the TELT platform, it is at this point that the systemic sub-layer cycle is triggered. The essential function of each sub-layer is to evaluate the:

1. educational value and usefulness of a technology application for inclusion in the TELT platform
2. learning effectiveness of approved TELT applications, and,
3. learning environments that are supported by approved TELT applications.

**L&T Sub-layer 1** evaluates the educational usefulness and value (the usability) of a technology application, tool or solution (juxtaposed with the Stage 3 business and technical layers). The purpose of sub-layer 1 is to avoid the risk of a widespread “technology driven pedagogy” mistake. Often faculty are advised to move “beyond intuitive value” and “adapt to the changing landscape” (Carmean and Brown, 2005, p. 5) in the implementation of new technologies, while the reasons behind the change such as technological progress or the commercial interests of LMS developers are not made clear. If the technology application addresses the criteria for all three (primary) layers, then it will be assessed for distribution to staff through the TELT Platform.

The premise on which sub-layer 1 rests is that technology is not merely an inert tool - it possesses inherent values (either in how it is designed or the way it functions) that may directly or indirectly influence pedagogical affordances. In other words, technology exerts its own pedagogical orientations (Britain and Liber, 1999).

**L&T Sub-layer 2** evaluates the learning effectiveness of an approved TELT application. This sub-layer is intricately related to the educational values of the instructor, school, faculty; learners’ characteristics; and, the learning environment. As such, sub-layer 2 is a crucial component of the iterative L&T sub-layer evaluation cycle in that is dependent on and also informs the evaluation methodologies and relevance of the criteria used in sub-layers 1 and 3. Highly graphic, visual snapshots of this type of evaluation can be found in the examples presented by Reeves and Laffey (1999). Although the authors have limited the scope of their evaluation to problem-based engineering courses, their evaluative approach is readily transferable to pedagogies and learning environments developed in other disciplines.

**L&T Sub-layer 3** evaluates the learning environments (the course design) in which TELT applications are employed. It incorporates many of the best practices and recommendations advanced by the educational technology field over the last decade (see Graham, Cagiltay, Lim, Crainer and Duffy, 2001 for an example). These approaches may be blended with or added to existing faculties’ and schools’ preferred design and delivery models and practices. In this way, the flexibility in methodology choice afforded by sub-layer 3 serves to enhance the learning effectiveness of online environments as well as inform the evaluative methodologies and relevance of the criteria applied in sub-layers 1 and 2 in the ongoing cycle of refinement and modification.

[Refer to Diagram 4]
The systematic / systemic structure of the framework and supporting evaluation methodologies are crucial to ensuring the long term success of the TELT platform as they underpin and are underpinned by the UNSW vision and guidelines, and evaluation methodologies identified for this framework as noted beforehand. It is especially critical that the evaluation methodologies should include the learner (students) in the evaluation process as opposed to confining them to the role of an end recipient of the delivered educational services, where all decisions are made without their input as in the case highlighted in Britain and Liber (2004). Equally vital, is the need to avoid a repetition of the case reported by Alexander and Golja (2007), where the learner is formally included in the evaluation process, but his/her voice is used to justify the adoption of the technology solution rather than making a genuine attempt to improve the learning outcomes.
References


Diagram 1

TELT EVALUATION FRAMEWORK

**Stage One:**
Vendor Assessment (1 to 2 days)
(determine suitability of vendor and/or open source solution)

**Stage Two:**
Pre-Pilot Scoresheet (2 to 4 weeks depending on scale of solution)
(determine suitability of technology for Stage Three evaluation)

**Stage Three:**
Full Pilot Evaluation (for length of pilot)
(determine suitability for TELT platform)
Stage Three of Evaluation Framework (expanded)

UNSW Vision and Principles + Teaching Approaches / Models
[as applied by Faculties and Schools]

Three layers of evaluation criteria

1. L&T Criteria
   
   Key Elements:
   
   • L&T criteria at university level
   • L&T criteria specific to faculties & schools
   • L&T Criteria relevant to technology application:
     • enterprise wide (Blackboard)
     • faculty / school choice of application suites
     • faculty / school specific applications (non-TELT)
   
   All informed by the L&T sub-layer evaluation cycle (Diagram 3)

2. Business Processes
   [informed by 1 first & also 3]

3. Technical Requirements
   [informed by 1 first & also 2]
Stage 3 Evaluation Criteria - L&T [expanded]

Three sub-layers of L&T evaluation

1. Stage Three of TELT platform (educational value and usefulness of selected application)
2. Evaluation of learning effectiveness of TELT technologies
3. Evaluation of learning environments that use TELT technologies (course design)
Diagram 4

Graduate Attributes
T&L Principles
UNSW 16 Guidelines

Teaching Approaches / Models
[key attributes]

- Stage Three Pedagogical Criteria (sub-layer 1)
- Technology learning effectiveness (sub-layer 2)
- TELT Learning environments (sub-layer 3)

Evaluation Methodologies
Evaluation of technology-enabled learning and teaching: establishing the need for a generic framework

1 Brief history of the educational technology evaluation debate

The evaluation of technologies in education has a long history. Starting from generic or industry-oriented evaluative models such as Kirkpatrick’s four steps (1983), the Xerox approach (Alden, 1983) and educational K-12 models (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman, and Provus, 1971), technology evaluation was primarily considered a subset of educational evaluation just as the virtual campus was mistakenly referred to as a sub-branch of the main (physical) campus during the initial stages of distance education expansion and experimentation.

However, similar to the manner in which the transportation revolution commenced through the invention of the automobile, technology turned out to be not just another tool or delivery method, but brought into existence a specific set of teaching and learning principles and curriculum requirements and thus established technology-enhanced teaching and learning as an educational domain in its own right, with its own rules and capable of delivering a wide variety of new outcomes for learners, lecturers and the educational institution.

One significant outcome, namely the influence of different instructional media on the learning process, was brought to the attention of pioneer educational technology evaluators as early as the 1990s. The learning <-> technology dichotomy generated considerable debate within the educational community with the central view that both aspects – the structuring of the learning activities (instructional design), and the methods for delivering those activities (technology) require careful evaluation in order to determine the totality of the learning process (Clark, 1994). Simply adapting an evaluation instrument used for traditional on-campus courses does not capture all the multiple facets of evaluating a technology-enhanced course (Roberts, Irani, Lundy and Telg, 2004). The notion of a learning <-> technology dichotomy exerted significant influence on the development of the educational technologies evaluation field, and is discussed further in the section titled ‘Biases’.

Taken as a whole, the optimistic predictions of educational psychologists on how technologies are best applied to support and enhance teaching and learning (Kozma, 1994), are still being validated today. A wide range of questions about the effective use of technologies in education still remain unanswered. The evaluative concerns expressed a decade ago by Bruce (1999) have only been partially integrated into current models. In the section to follow, the discussion will focus on the questions of why the evaluation of educational technologies presents such a challenge for educational communities and what can be done in order to resolve the complex question of how to devise an all-inclusive framework for the evaluation of educational technologies.
2 Advantages and drawbacks of existing evaluative frameworks

There are a variety of evaluation models that are used in the educational technology field. They range from toolkits (Oliver, McBean, Conole and Harvey, 2002) to cookbooks (Learning Technology Dissemination Initiative, 1998), along with new evaluative models that are still under development such as the institutionalisation model of Nelson, Post and Bickel (2001) or the Software Process Improvement and Capability dEtermination (SPICE) model (Marshall and Mitchell, 2004). The questions of why new models are still appearing after twenty or so years of educational technology research and what are the contributing reasons raises a number of issues in reviewing the literature on educational technologies evaluation:

2.1 Technologies are in constant development

Within the domain of computer science, the word “iteration” has a specific meaning: any project progresses through several iterations; any new tool is refined through several development-test-improvement cycles; and, the development of new tools commences with prototyping. Such iterative processes mandates the need for an evaluative framework that is also iterative in its approach to help solve the “once something is completed it is already obsolete” (Bruce, 1999) dilemma encountered by many researchers with an educational, curriculum, and instruction background.

Traditional formative and summative evaluation methods do not neatly fit the eternal kaleidoscope of changing technological tools, systems, and architectures. Although formative evaluation aims to assess products and processes as they are developed, the technological development cycle never ceases. Thus, in the sense of establishing a defined end point for a completed project or product, the inherent goal of summative evaluation may never be feasible. The development process is subject to transition from one stage of its development to the next. Therefore, in the area of educational technology evaluation, there is a cogent need for an evaluation framework that accommodates the iterative evaluation processes that are characteristic of many technology-related processes, whilst as argued by McNaught (2002), simultaneously embracing the eclectic myriad of teaching and learning functions required for successful delivery through the Web.

To some extent, these problems have been addressed by several evaluation models. The Marshall and Mitchell model (2004) includes iterative evaluation cycles; and the Agostinho et al., (2002) Evaluation and Redevelopment Framework (ERF) model includes possibilities for addressing the technological redevelopment issues. However, both these models, one being technological, the other embracing concerns about the needs of learners, do not address the relevance of the applied pedagogies, or detail the adequacy of the technological evaluation criteria. The first model is discussed further in the Biases section, while the second is expounded on in the section titled ‘A Generic Evaluative Framework’.

2.2 Narrow scope
From the times when eLearning, technology-enhanced learning, and distance education were viewed as just the other delivery option within the broader university context, the idea of educational technologies evaluation morphed into a programme of evaluation, something that is conducted along administrative lines and is narrowly focused on the institution’s preparedness to support and promote the program. By starting with an industry-oriented, business-related evaluative model, the educational technologies evaluation field could not escape one of the fallacies of a business-related approach when the scope of evaluative activities is limited to a specific delivery programme (Agostinho et al., 2002; Franklin, Yoakam, and Warren, 1996), or even a single tool (Crozat, Hû, and Trigano, 1999). According to Lefoe, Gunn and Hedberg (2002), this tendency translates into evaluation reports that are relevant only or are confidential to the particular institution (or country) that initiated the study.

In general terms, as Baker and Herman (2003) note, two prevalent approaches to technology evaluation in the past fifteen years are: the evaluation of specific technologies (a narrow scope), and broad, often longitudinal, studies where the results become irrelevant long before they are realised due to the rapid rate of change in technology. Indeed, even models with a clearly articulated pedagogical agenda such as the Britain and Liber (2004) conversational model focus on existing Course Management Systems in their evaluative purposes, but avoid examination of broader enterprise level solutions or virtual learning environments such as Second Life (2010).

The essential problem with a narrow evaluative scope as applied to technology evaluation is that information technologies are also communication technologies, and thus produce a network effect (Bruce, 1999). A tool used in one classroom may be used very differently in another context; all collectively generating a relatively broad application area that may or may not be visible in the framework of one institution or programme.

2.3 Theory base

While early educational technology evaluation models were grounded in psychological / measurement theories, the rapid rate of change in technology generated a situation where a planned evaluation was not completely thought through at the implementation or even the earlier, programme design phase. Out of necessity, the evaluation study was conducted ad hoc by people with varying, non-evaluative backgrounds in an attempt to assess / improve an existing technology (this point is discussed further in the Biases section).

As a result, there are a number of evaluative models that do not contain a strong rationale, theoretical foundation, or a statement of purpose, and attempt to compensate for the absence of these elements by; presenting evaluators with a context-specific selections of questions to submit to interviewees (Mehlenbacher, Bennett, Bird, Ivey, Lucas, Morton, and Whitman 2005); or advancing a “just-in-time” evaluative delivery approach (Reeves and Hedberg, 2003); or preparing tailor-made evaluative plans that draw on an existing pool of questions (Lam and McNaught, 2004). Other models (such as the Curriculum Corporation, 2004 model) produce a self-fulfilling effect in the sense that the evaluative criteria or
schema are not connected to existing literature, but instead rely on internal design guidelines.

2.4 Prescriptive vs. descriptive models

In general, evaluation models can be divided into prescriptive (most common), which specify how evaluations are conducted (exemplars); or descriptive, a set of statements and generalisations that describe, predict, or explain evaluation activities. Prescriptive models may include iterative evaluation cycles (Marshall and Mitchell, 2004), but in essence the iterations are imposed on evaluators and stakeholders instead of being an integral part of the evaluative activities.

Descriptive models, on the other hand, allow for an authentic reflection of the evaluation activities in real time (Pangaro, 2000) and can be considered a viable solution in the case of educational technologies because of the rapid rate of technological change as outlined beforehand. Furthermore, descriptive models are participatory: the stakeholders themselves are involved in an empowering process of democratic decision-making and the active promotion of social change. These models better reflect the reactions and impressions of the participants and stakeholders. In such instances, where the evaluation is conducted in a not-for-profit organisation (for example, a public educational institution), the descriptive evaluation model presents a viable alternative to a prescriptive model. One other reason to consider a descriptive, participatory evaluation model is the fact that all the stakeholders (including students) are given control over the creation of the knowledge (Cousins and Whitemore, 1998) and the informational input needed to effect the evolution of a final summary.

However, within the educational technologies evaluation field many models appear to be prescriptive, with rare exceptions (Lam and McNaught, 2004). However, even these exceptions represent a collection of evaluation-related activities rather than full-scale models, and simply offer basic background information on how to evaluate different pedagogical, technological and organisational components.

2.5 Value-free vs. multiple values evaluation

The notion of value-free evaluation is closely related to the questions surrounding evaluator-driven (prescriptive) versus stakeholders and evaluator driven, real time descriptive evaluation. The prescriptive approach inevitably generates the question of “whose values” are reflected in the particular evaluation process. One way to solve this problem is to embed multiple values in the evaluative process (Zammuto, 1982).

Embedding multiple values in an evaluative framework allows for avoidance of the evaluator's biases as well as assists to reveal a detailed picture of the environment / technology under examination. This strategy also requires more resources and time to be allocated to the evaluative activities (Chen, 1990). However, modern pluralistic societies consider addressing the needs of different members of the evaluation audience is a
necessary attribute (Stufflebeam and Shinkfield, 2007). While there is little doubt about the necessity, it is also a difficult task to resolve.

Although there is no shortage of evaluative frameworks or evaluation models for different types of technology, there is however, a shortage of generic, inclusive, multiple-values evaluations of educational technologies that are grounded in a teaching and learning philosophy. Existing models reflect the values of the evaluator, institution or even a software manufacturer (as highlighted by comparing the Institute for Higher Education Policy benchmarks (2000) and Marshall’s eMM (2007), which is detailed in Biases (to follow). Recent evaluative models are still struggling with the multiple values challenge (Lam and McNaught, 2004) and the prescriptive versus descriptive challenge (Bach, 1994).

3 Biases

Following directly on from the previous section, the research literature reviewed clearly confirms that the learning <-> technology dichotomy was widely discussed by the educational technology community during the 1990s. This dichotomy also influenced the field of technology evaluation since the evaluators themselves were either from a software engineering background, or a pedagogical background, and many were not professional evaluators. These two major backgrounds (technological and pedagogical) generate two distinct lines of inquiry in examining the purposes and developing evaluation models. While some models espouse a technological stance (Marshall and Mitchell, 2004; Mehlenbacher, et al., 2005), other models support a pedagogical approach with a minimal attention paid to organisational and technological details (Boud and Prosser, 2002; Dean, Biner, and Coenen, 1996).

In particular, Marshall and Mitchell’s SPICE includes a learning / pedagogical component in its later version, but the component itself (that is, the questionnaire) has been updated from the study sponsored by a major educational software company. Thus, a certain bias favouring this particular company may be present in the latest version of SPICE. On the other hand, Boud and Prosser’s model highlights technology features that are important for the development of learners’ critical thinking abilities and do not cover the crucial organisational / implementation aspects of technology evaluation.

While there are models that blend technological and pedagogical approaches, they have not escaped the challenges inherent in evaluating technological factors such as the innate structure of the application that determine interface design and navigational capacity, and the impact of these aspects on learning (Agostinho, et al., 2002). In some instances, the organisational questionnaires used in these models are far more detailed than the pedagogical component (Attwell, 2006, tools for SME).

Overall, educational technology evaluation requires new integrative approaches that avoid the faulty $Evaluator = Evaluation$ formula that is independent of the evaluator’s background, and covers multiple aspects of the evaluation process starting from the organisational to the pedagogical, and extending to socio-cultural factors.
A substantive reason a pure technological approach to educational technology evaluation became popular is the effect of technological progress on the growing “informatisation” of society, that is, the “shift of the central sources of economic activity from the primary and secondary sectors towards the services and information sectors of the economy” (Cunningham, Ryan, Stedman, Tapsall, Bagdon, Flew, and Coaldrake, 1998, p.13). Technology is in vogue, and is no longer viewed merely as a tool - it possesses integral values that may directly or indirectly influence pedagogical affordances and uses. Often, the fact that technology has its own pedagogical orientation (Britain and Liber, 1999) is totally overlooked.

Faculties are advised to move “beyond intuitive value” and “adapt to the changing landscape” (Carmean and Brown, 2005, p. 5) of new technological implementations while the reasons that makes the landscape change – technological progress or the commercial interests of software developers, are still unclear. It is for this reason that viewing a LMS (for example) as a starting point for learning in online environments was widely criticised in the research literature (Siemens, 2004).

Meanwhile, one notable approach to educational technology evaluation gave rise to beneficial outcomes, namely quality assurance standards and benchmarks for technology developers and users. To some extent, quality assurance benchmarks in distance learning and education promoted the development of technology evaluation by creating structural elements on which to direct and guide the evaluation process. However, benchmarks should be developed systematically, not ad hoc. The early examples of “cherry-picking” best practices in education and devising benchmarks based on those practices failed to live up the promise (Hagner, 2001).

It is also necessary to distinguish evaluation benchmarks from evaluation models. By definition, benchmarks provide a fixed baseline against which to gauge the progress and effectiveness of the evaluation process as opposed to the much broader evaluation model (or framework) that assists to structure the evaluation process to explore the multiple facets of the phenomenon under examination.

What happens if the baseline is skewed (or even biased)? The Institute for Higher Education Policy benchmarks (2000) were widely adopted for technology evaluation. The benchmarks generated by this study (which included a set of Institutional support, Course development, Teaching and Learning, Course structure, Student support, Faculty support, and Evaluation and assessment guidelines) were further adopted by Marshall for his eMM model (2007)* and referenced by many educational technology researchers. The problem however, was that this set of benchmarks was originally generated by one of the largest monopolies in the educational software field (the study was commissioned by Blackboard Inc.). Although Blackboard was not a monopoly in 2000, the ability to describe and influence the major functions of their software tools during the development phases may have aided in the transition to monopoly status. The potential to guide the outcomes of software development

---

* Marshall included the same benchmark statements under different categories: Learning, Development, Co-ordination, Evaluation and Organization.
may prevent other important needs from being met by educational software companies, particularly if a widely accepted of benchmarks was already in place.

Then there are instances where the practices and the processes of teaching and learning are largely ignored in the pursuit of functions, features, integration, and a myriad of technological and business concerns. Thus, quality assurance benchmarks are useful in reflecting the essential baseline, but the origin of those benchmarks should also be carefully examined in order to prevent a situation where the development of the evaluation criteria is conducted by the entity that in effect, is the subject of the evaluation.

One other bias related to the issues noted above is what could be referred to as the 'additive' approach to technological integration. When different technology systems are integrated (such as BlackBoard and student record systems), the question of how teaching and learning will benefit is often overlooked (Boys, 2002). This situation is not simply the result of implementation issues, but is a direct consequence of whenever faculty voices are not heard or not taken into account by educational administrators. Moreover, although there are instances of benchmarks related to students, there is no evidence of attempts to refine such benchmarks to include them as vital stakeholders in the evaluation and modification of educational technology tools.

The integration of different technologies should promote the use of different learning and teaching philosophies and methods. Thus, a generic evaluative framework that facilitates evaluation of the final product or the synthesis of several technological systems may prove to be a necessary alternative to the piecemeal evaluation of different components of a complex technological mosaic. Britain and Liber (2004) also argue for a comprehensive (learner-centred) approach to evaluation as opposed to an additive (administration-centred) approach.

In their study, Britain and Liber confirmed that virtual learning environments changed markedly during the five years from 1999 to 2004 in that they focused less on teachers and administrators, and more on learners themselves. Nevertheless, they believe that inherent biases still exist since the largest educational software manufacturers still use third party products (for example Oracle) to improve the quality of their software offerings. Thus, ultimately virtual learning environments design could also be considered technology (third party) driven.

It is feasible that unwanted biases could be eliminated with the formulation of a generic evaluative framework that encompasses research findings from completed educational technology evaluations. Such a generic framework should incorporate a wide range of teaching and learning practices that are not confined to one particular institution or by the outcomes of one specific sample. This framework needs to be broad in its scope to inform and improve learning and teaching practices conducted within and across different departments, schools, faculties, and institutions.

Last but not least, in addition to enhancing learning and teaching, the evaluation framework should also assess how well the theoretical framework itself improves technology
evaluation activities and informs how best to incorporate new technology tools into teaching and learning practice. The long-term goal of the TELT evaluation strategy is to devise an evaluation framework that encompasses the iterative processes that underpin technology development, tracks the trajectory of success across faculties, and over time enhances divergent learning and teaching approaches.

4 Open Source evaluation

The evaluation of Open Source solutions stands out as a separate area from general (commercial) technology evaluation models. Partially, it can be explained by the origin of the Open Source software philosophy as discussed below; and partially by the desire of the commercial software companies to eliminate free educational technology options by integrating Open Source products into their own products range.

In 2005, Barbara Ross (a co-founder of WebCT) envisioned the integration of the open source and commercial tools in an attempt to support diverse pedagogies and learning outcomes (Ross, 2005). According to Ross, the main features of such integrated systems should include support for teaching, learning and research. However, this approach appeared to view open source software as no more than another piece of cake on the institutional enterprise plate. Although on the surface, the idea of integration is worthwhile and thought-provoking, no organisational or technical details followed from the vision statement made by the author of the chapter.

An alternative integrative approach was proposed by the educational software creator - ANGEL Learning. In their view, course management systems should be open to allow low-level customisation with the client institution’s information technology services and needs because it is not always cost-effective for the educational software creators to embed this level of customisation (Mills, 2005). Is open source just a quick and cheap solution for the customisation problem or is it a stand-alone instructional solution? One of the best features of Open Source platforms that made them popular with Asian distance education institutions is exactly the ability to handle customisation issues in line with the needs of developing countries (Batpurev, 2005).

The preferred model for educational technologies evaluation should acknowledge the uniqueness of the conditions that have brought Open Source into existence. First, according to an early theorist of the Open Source movement Eric Raymond, Open Source originally came into play out of necessity to “scratch a developer’s personal itch” and the strong personal motivations of the developers involved into Open Source projects that almost guarantees high quality software solutions (Raymond, 1998). Second, software that is open to everyone, such as Facebook or a blog, tend to prompt users to cooperate and apply their efforts to attainment of a self-regulated learning ethos. Moreover, open source software presents a less threatening learning environment for students when compared to the reduced flexibilities and restrictions of commercial products.

In a sense, commercial products emulate the traditional educational system with their inherent structures and divisions that may serve as a conscientisation tool (Freire, 2007)
that manoeuvre students to inadvertently absorb the boundaries and functional order that in turn may intimidate or disrupt the learning effectiveness experienced by other more independent, constructivist orientated students. By default, most commercial learning management systems afford an objective / instructivist course design approach, whereas open source solutions permit constructivist learning approaches.

In a Freirian sense, Open Source presents a more “authentic” approach to learning thus allowing students and developers to be aware of their individual learning needs as well as the shortcomings of the software and as a result, seek ways to work cooperatively to solve these issues. The collaborative opportunities afforded by Open Source software assist to invert imposed limitations and allow truly democratic concepts of education and technological literacy to arise (Stokes and Stokes, 1996).

The potential for Open Source platforms to add pedagogical / critical theory value to the learning process clearly demarks the core difference between Open Source and proprietary Course Management Systems (CMS). The latter according to Arnold (2003), appeared to have cut training costs, and their raison d’être has a business rather than educational nature. Thus, two highly contradictory tendencies reflected in Open Source software solutions can be identified: it is software developer-driven; and at the same time, presents a less restrictive learning environment for students. It also illustrates the innate influences of applying a pedagogical versus a technologically led focus to developing (and evaluating) educational technology.

Unfortunately, most literature on Open source evaluation is limited to technological and business factors without taking the pedagogical characteristics of the software into consideration. For example, the Karin van den Berg synthesised model (2006) includes such factors as community, release, activity, longevity, license, support, documentation, security, functionality, and integration. However, her notion of community refers to software developers and technologically savvy audience rather than educators. In a similar way, several additional factors in van den Berg’s model refer to technological and business variables. A further example of a technological approach to Open source Evaluation is presented in the work of Graf and List (2005) where all pedagogical factors are covered under the single checklist category of ‘user-friendliness’.

Nonetheless, despite the variety and sophistication of evaluation approaches that can be applied Open Source software, the positive effect of Open Source concept on students’ learning appears to have been unnoticed by the educational technology community.

5 Attributes of a Generic Evaluation Framework

5.1 Methodologies

One of the first steps toward the creation of the generic evaluative framework was made by Hedberg, Wills, Oliver, Harper and Agostinho (2002). They revised the possibilities for developing generic evaluation rubrics based on high quality learning activity designs. Their model included the key elements of learner’s engagement, learning context, challenge and
practice. However, their approach, just as in the case of Britain and Liber (2004), seem to focus on a “micro-world” / learning activity level and used a grounded theory methodological approach where existing practices “build” the new evaluation framework. This approach holds promise for existing technologies, but it neglects the iterative nature of new tools as they become available in an ever changing world of technological advancement. As with any grounded theory case, it perfectly identifies what is already in place, but does not cover crucial additional practices and activities that may not have been included in the sample. It is widely accepted that the main purpose of an evaluation study often defines and therefore limits its scope (Learning Technology Dissemination Initiative, 1998).

In order to avoid the mistakes of previous models, in particular, limiting the evaluation sample and using a grounded theory approach, Agostinho, et al., (2002) emphasised the importance of choosing an appropriate evaluation methodology (or several methodologies). To be truly generic however, it is argued that methodological considerations should incorporate a flexible system of evaluating different qualities and attributes similar to Marshall’s (2007) model, in which each question is assessed in terms of the adequacy of its representation in a particular context. An extended review of evaluative methodologies is presented in the document titled “Review of Methodologies of Evaluating Technology-Enabled Learning and Teaching” (Appendix 1).

As noted several times, students should also be considered valid stakeholders, and so one of the elements of the evaluative methodology should permit full expression of their concerns and encourage the right to influence all recommendations to adopt a particular educational technology. Alexander and Golja (2007) actively included students in a process of questionnaire refinement, but students were not considered decision-makers in the selection of different forms of technology. Instead the students’ voices were used to map potential problems with existing systems and resources. A methodology allowing for automatic inclusion of faculty, students’, IT experts’ and administrators’ voices holds genuine promise, not only for resolving potential methodological problems, but also for instantiating a multiple-values evaluation approach as noted beforehand.

It is important to note that the timing of evaluative activities (before, during or after a certain technology is implemented) is also intricately related to the selected methodology and the format of the evaluation used (refer to Tables 1 and 2 for an example). Owen (2006) recommends several methodologies and methods for the accumulation of evidence that may improve the final research findings, which in turn may better inform subsequent iterative cycles.

5.2 Formative and summative aspects, culture of evidence

From the outset of the technology evaluation era, a precedent was more or less been established when it is considered that less than fifty percent of evaluative studies were based on formative activities (Alexander and Hedberg, 1994). As noted earlier, the rapid rate of change in technology often negates the relevance and meaning of applying timeframes to summative evaluation studies. An iterative formative evaluation process not
only allows for a review of technological change, but also to establish a solid body of evidence from which to resolve the following question: What steps can institutions take to assist faculty members and other staff to gather and use evidence in order to improve learning and teaching practice?

The broader question of the scholarship of learning and teaching may involve the question of the number of participants/adopters as the more that become involved in enhancing their practice using technology, the greater an asset the TELT platform becomes for the university. As observed in education generally, the core conceptions of learning and teaching have changed “from imparting and acquiring ‘content’ to facilitating and constructing knowledge, a shift from ‘product’ to ‘process’” (Cunningham et al., 1998, p. 32). Thus, an iterative formative approach may serve to address the evaluation requirements.

A generic evaluation framework should also allow for several evaluative outcomes that include a capacity to reflect on the educational value and effectiveness of the framework itself. That is, in order that it can contribute to improving learning effectiveness, the TELT evaluation framework should be evaluated to determine the degree to which such outcomes were actually achieved. Once again, an iterative evaluation process is required to ensure that specific activities may be modified in line with the results as they become available. Moreover, since all evaluation results will be disseminated, and the merit of the new framework is based in part on the evaluation outcomes, the framework also requires summative evaluation strategies to allow for feedback and comment, and so ensure a balanced approach is consistently applied to selection and use of educational technology.

5.3 Major components

The effective alignment of a new evaluative framework to present and future learning and teaching needs requires that it draws on previous research in the field. As such, pedagogical, organisational and technical issues are generally viewed as the core themes in many evaluation studies (Conole, 2004). However, from a European perspective, Attwell’s (2006) report on the use of management models for evaluation (a business component) and the lack of integrative models that take into account both technology (iterations) and learning processes, validates the concerns that initially led to this review. That is, the report confirms that policy evaluation is a vital aspect of the overall evaluative process.

Since the evaluation of technology-enhanced learning is often regarded as “the conflict between modern, positivist, experimental and postmodern, critical, case-based philosophies” (Gunn, 1999, p. 186), and given the mandate to enhance learning through the use of technology, it is more productive not to enter into such philosophical debates and instead ensure that all TELT platform policy, technical and organisational factors are guided primarily by pedagogy. At a broader level, the pedagogical component will need to observe a comprehensive range of questions starting from “How does the learning environment allow the lecturer to present a concept?”
One of the assumptions of existing evaluation frameworks that include a pedagogical component is that institutional policies and organisational structures are already in place for faculty and learners. Both the Boud and Prosser (2002) and Dean, Biner and Coenen (1996) studies are based on this perception. However, this position does not always hold true and therefore the business processes and administrative components should not be neglected in defining a generic evaluation model.

Thus, the framework should also focus on the managerial (organisational), strategic goals of evaluation as well as the essential business processes that together underscore the overall effectiveness of the evaluation process. The selection and use of ICT developments that support learning and teaching are no longer predominantly confined to isolated projects within academic departments and learning technology support units. Instead, ICT should form a core part of institutional policy (Conole 2004).

Other useful findings from the literature review include the need to consider the socio-cultural factors and governmental agendas that sit closely behind the adoption and use of technologies (Conole, 2004); the concept of lifelong learning (Franklin, Yoakam, and Warren, 1996); and the characteristics of lecturers’ (Dean, Biner, and Coenen, 1996), that in part determine the key attributes that assist to establish the evaluation criteria for the pedagogical component.

As demonstrated by Britain and Liber (2004), and noted earlier, the generic framework should also engage students in the evaluation process rather than viewing them as recipients of an educational service where all decisions are made on their behalf. The new framework should view the learner as self-regulated participant, responsible for his / her own learning as a collaborator and peer in an ongoing learning and evaluation process (Chickering and Ehrmann, 1996).

The business processes component requires a comprehensive review of attributes such as meeting the demands for new job skills; cost control; the provision of widespread access to training/education opportunities; attracting/engaging new learners; and building an adequate labour force for community businesses. The business processes component also needs to include organisational and structural issues. In some cases, institutional strategy is “the sum of large numbers of independent actions by many faculty members and students across the college” (Ehrmann, 1995). The organisational structure or a planned structure can often turn out to be quite different from the projected structure, and so if educational administrators did not plan for such inconsistencies then evaluation of administrative activities in relation to learning and teaching and faculty development can serve identify crucial gaps.

Another advantage of a generic framework is that it could assist not only to improve organisational processes, but also resolve issues of organizational change that are often ignored in an additive approach to technologies evaluation. A detailed schema of business processes evaluation could reciprocate and put in place an explicit model for the management of change, relevant to tertiary education (Boys, 2002). Traditional evaluation
models often do not explain why changes occur, how changes are different across settings, or how they relate to changes in the innovation (Bruce and Rubin, 1993).

The third evaluation component, technical, has been relatively well-covered in previous research studies such as Obexer’s model (2005) and the Curriculum Corporation model (2004). Whilst covering technological factors, the Obexer model provides a narrow scope for evaluation in terms of including tools that are comparable to learning management systems, but the model does not allow for inclusion or iterative analysis of faculty specific tools and practices.

The Curriculum Corporation (2004) framework incorporates three distinct layers of evaluation and takes into consideration business as well as technological processes. However, the pedagogical layer is relatively inadequate, not learner-centred, emphasises mainly the issues of security on the Web, and presents student engagement as a unidimensional process (that is, yes or no). Moreover, the framework is aimed at evaluation of one particular product with some comparative attempts made. It does not position a product within a broader array of technological tools to enhance learning and teaching. A generic framework will therefore aim to expand the technological component to include the iterative technology-development and feedback cycles. It must also ensure that all three evaluation components directly reference the core attributes of learning and teaching scholarship as a primary focus.

The final point to make is the necessity to include “worldware” (Ehrmann, 1995) such as email, internet, and word processors in a generic evaluative framework. All software is not particularly designed for learning such as an LMS, but often serves an important function in learning and teaching processes that should also be the subject of detailed evaluation.

6 Integrative models and their shortcomings

Over time, there have been several large-scale attempts to create an integrative evaluation model for educational technology evaluation. The European project (Attwell, 2006) referred to earlier can be considered one significant example. This project however, presents an eclectic collection of evaluative frameworks that are applicable to the educational / geopolitical contexts of various European countries. No generic or even comprehensive approach that is predominantly focused on learning and teaching, or technology enabled teaching and learning evaluation, can be identified in this collection.

One other attempt however, the Benvic Benchmarking system (2000, 2002) definitely extends beyond the narrow scope of a context-specific benchmarking methodology. It satisfies the conditions of a generic framework in presenting multiple pathways to technology evaluation, but recommendations made are based directly on the findings of previous field research and as a result presents a sound theoretical base. Nevertheless, the model is too abstract for practical application. There are no concrete guidelines, or a variety of evaluation methods or ready-to-use checklists to be found in the Benvic deliverables document (Benvic Project, 2000).
The later report (Benvic Project, 2002) covers administrative/infrastructure questions, but the pedagogical or technological variables are too broadly defined with little attention paid to providing details on the relevant pedagogies. A similar issue occurs with many large scale evaluative models that again give rise to questions that are too broad in their scope to resolve the development useful pedagogical and technical criteria, while overemphasising the institutional structure factors (McNaught, Phillips, Rossiter, and Winn, 2000).

7 Summary Conclusions

In summarising the outcomes of this review, consideration has been given to the variety of technological tools employed by educational institutions and the wide range of uses of those tools as determined by the institutional context. At this stage, the present report provides an overview of educational technology evaluation models and identifies the importance of certain factors that have been omitted in the evaluative models that were the subject of this review.

Based on the review to this point (which is also subject to further analysis and refinement), a set of recommendations is presented as a basis for formulating a generic TELT Evaluation Framework that aims to afford an in-depth examination of how to apply established learning and teaching principles to divergent contexts using selected educational technologies to enhance the practical day-to-day application of these principles. For now, the key recommendations include:

1) an iterative, formative evaluation cycle (a feedback-improvement cycle)
2) avoid a narrow evaluative scope, but facilitate “scaling down” to the especial needs and of particular departments or schools
3) synthesise the major literature findings regarding the role of pedagogical, technological and business variables as applied to the TELT platform evaluation process
4) apply a descriptive, participatory evaluation approach
5) develop a fully inclusive, multiple stakeholders and multiple-values evaluation model where evaluation forms directly align to the timing of an evaluative activity
6) employ multiple methodologies to assess each technological element and integrate learners’ participation into evaluative process.
7) assemble a core set of educational technology tools and “worldware”.

alignment with pages 2 – 4 advantages and drawbacks?
<table>
<thead>
<tr>
<th>Orientation</th>
<th>Proactive</th>
<th>Clarificative</th>
<th>Development</th>
<th>Monitoring</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synthesis</td>
<td>Clarification</td>
<td>Improvement</td>
<td>Checking/Refining</td>
<td>Learning / Accountability</td>
</tr>
<tr>
<td>Typical Issues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is there a need for the program?</td>
<td></td>
<td></td>
<td></td>
<td>Has the program been implemented as planned?</td>
</tr>
<tr>
<td></td>
<td>What do we know about this problem that the program will address?</td>
<td></td>
<td></td>
<td></td>
<td>Have the stated goals of the program been achieved?</td>
</tr>
<tr>
<td></td>
<td>What is recognised as best practice in this area?</td>
<td></td>
<td></td>
<td></td>
<td>Have the needs of those served by the program been achieved?</td>
</tr>
<tr>
<td></td>
<td>Have there been other attempts to find solutions to this problem?</td>
<td></td>
<td></td>
<td></td>
<td>What are the unintended outcomes?</td>
</tr>
<tr>
<td></td>
<td>What does the relevant research or conventional wisdom tell us about this problem??</td>
<td></td>
<td></td>
<td></td>
<td>Does the implementation strategy lead to intended outcomes?</td>
</tr>
<tr>
<td></td>
<td>What do we know about the problem that the program will address?</td>
<td></td>
<td></td>
<td></td>
<td>How do differences in implementation affect program outcomes?</td>
</tr>
<tr>
<td></td>
<td>What could we find out from external sources to rejuvenate an existing policy or program?</td>
<td></td>
<td></td>
<td></td>
<td>Is the program more effective for some participants than for others?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Has the program been cost-effective?</td>
</tr>
</tbody>
</table>

|----------------|------------------|--------------------|--------------------------|-------------------------|-------------------|---------|------------|-----------------|----------------|-------------|-----------------|---------------------|--------------------------|----------------|------------------|-------------|-----------|-----------------|----------|-----------------|

* Both tables are reproduced with permission from the author (Owen, 2006, pp. 52-53).
### Table 2. Forms of Evaluative Enquiry

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Proactive</th>
<th>Clarificative</th>
<th>Development</th>
<th>Monitoring</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Issues</td>
<td>Synthesis</td>
<td>Clarification</td>
<td>Improvement</td>
<td>Checking/Refining</td>
<td>Learning / Accountability</td>
</tr>
<tr>
<td>State of Program</td>
<td>[see Table 1]</td>
<td>[see Table 1]</td>
<td>[see Table 1]</td>
<td>[see Table 1]</td>
<td>[see Table 1]</td>
</tr>
<tr>
<td>Major Focus</td>
<td>None</td>
<td>Development</td>
<td>Development</td>
<td>Settled</td>
<td>Settled</td>
</tr>
<tr>
<td>Timing</td>
<td>Program Context</td>
<td>All Elements</td>
<td>Delivery</td>
<td>Delivery/Outcomes</td>
<td>Delivery/Outcomes</td>
</tr>
<tr>
<td>[vis-a-vis Program Delivery]</td>
<td>Before</td>
<td>During</td>
<td>Mainly During but could be At Other Times</td>
<td>During</td>
<td>After</td>
</tr>
</tbody>
</table>

#### Key Approaches
- Needs Assessment
- Research Synthesis
- [Evidence Based Practice]
- Review of Best Practice
- [Benchmarking]
- Evaluability Assessment
- Logic Development
- Ex-ante
- Responsive
- Action Research
- Developmental
- Empowerment
- Quality Review
- Component Analysis
- Devolved Performance Assessment
- Systems Analysis
- Objectives Based
- Needs Based
- Goal Free
- Process-Outcome
- Realistic
- Performance Audit

#### Assembly of Evidence
- Review of documents and data bases, site visits and other interactive methods. Focus groups, nominal groups and delphi technique useful for needs assessments
- Generally relies on combination of document analysis, interview and observation. Findings include program plan and implications for organisation. Can lead to improved morale
- Relies on intensive onsite studies, including observation, degree of data structure depends on approach. May involve providers and program participants.
- Systems approach requires availability of Management Information Systems [MIS], the use of indicators, and the meaningful use of performance information
- Traditionally required use of pre-ordinate research designs, where possible the use of treatment and control groups, and the use of tests and other quantitative data. Studies of implementation generally require observational data. Determining all the outcomes requires use of more exploratory methods and the use of qualitative evidence.
References


Hypermedia & Telecommunications (pp. 729-735). Denver, Colorado, USA: Association for the Advancement of Computing in Education.


Appendix 1:

REVIEW OF METHODOLOGIES FOR EVALUATING TECHNOLOGY-ENABLED LEARNING AND TEACHING

This paper presents a summary of the findings to emerge from an exploration of studies on methodologies for evaluating technology-supported learning and teaching that may inform the development of suitable evaluation criteria that potentially may be incorporated into Stage 3 of the TELT Evaluation Framework (Full Pilot Evaluation). This stage determines the suitability of eLearning applications or solutions for inclusion in the TELT platform.

Stage 3 consists of three layers: Pedagogy, Technical and Business. The Pedagogy Layer is emphasised first and foremost as it is directly informed by the UNSW Graduate Attributes, Learning and Teaching Principles and the 16 Guidelines for Learning and Teaching Practices. The evaluation criteria specified for this layer are also guided by the teaching approaches and models applied in faculties and schools. The pedagogical criteria are further divided into three sub-layers: (1) the educational value and usefulness of the selected application, (2) the learning effectiveness of the TELT technologies, and (3) the learning environments that use TELT technologies. This paper is structured to accord with the proposed Stage 3 of the TELT Evaluation Framework and the themes that emerged from the findings of the literature review and analysis. Sub-stages 2 and 3 are conducted after the Stage 3 sub-layer 1 evaluation process is complete and the TELT application is released to the university community.

The educational technology evaluation methodologies literature were reviewed and summarised across the following criteria:

- Name of the methodology and its creator(s)
- Evaluation methods involved
- Instruments described in the documents
- Relevance to the TELT Evaluation Framework

Five categories of evaluation methodologies were identified in the course of completing this review. Although not directly relevant to purpose of this paper, these categories will serve as a useful guide in selecting evaluation methodologies once the framework is operational. The identified categories include:

- Fully-developed solution: includes rationale, procedure, ready for use instrument
- Semi-developed solution: a list or a system of evaluation criteria and questions.
- Theoretical framework: theories and principles that assist to develop evaluation methodologies for specific contexts.
- Meta-methodology: methodologies used to develop evaluation methodologies for specific subject matter domains.
- Evaluation guide: generic guidelines and templates of various evaluation methods.
1 The Pedagogy Layer

1.1 Educational Value and Usefulness of the Selected Application

1.1.1 Overview

A software application selected for pilot testing on the TELT platform is evaluated on the basis of its educational value and usefulness for learning and teaching. For the purpose of this paper, three important terms are defined as:

- **Educational value**: a term to indicate the integration of knowledge in the learner (that is, the transmission of knowledge to the learner) and the integrated development of knowledge by the learner (Crittenden, 1968)
- **Usefulness**: a term that applies to the capacity to apply pedagogical strategies to effect a positive or improved learning outcome (that is, learning effectiveness).
- **Usability**: a term to describe the ease to which users can perform tasks on an eLearning software application or solution to achieve particular learning goals associated with those tasks.

A detailed discussion on educational value and (learning) effectiveness is taken up in section 1.1.7. Although the terms usefulness and usability appear similar, there are several distinctions that are observed throughout this document:

- Usefulness informs educational rationales, while usability informs the usage of the software per se, that is, the usage may or may not reflect educational purposes
- The measurement of usefulness requires teachers’ and students’ participation whereas usability may be evaluated by experts who may not be educators.
- If the eLearning application is not designed according to educational principles, its educational usefulness is low, but the usability could still be high as long as the users feel it easy to perform the tasks they are supposed to perform.

Notwithstanding the preceding distinctions, research shows that the usability of eLearning applications has a significant positive correlation with students’ learning outcomes, which in turn reflects the educational value (Meiselwitz and Sadera, 2006). There are many usability evaluation methods available in literature that in general terms can be classified into three categories (Hom, 1998):

- Inquiry (field observation, questionnaire, logs, etc.)
- Inspection (heuristic evaluation, cognitive walkthrough, etc.)
- Testing (thinking aloud protocol, co-discovery learning, etc.)

Bowman, Gabbard, and Hix (2002) reviewed a range of literature about eLearning evaluation, summarising the methodologies according to the user’s requirements, evaluation context and the types of results, from which they developed a classification model (Figure 1).
Bowman et al.’s (2002) model is useful for methodology selection to accord with varying conditions (for example, stage of programme development; personnel availability; usability dimension to be measured; whether quantitative results are needed; and whether remote evaluation is needed). Selection of a method is dependent on the particular context in which the application is applied. The criteria for most usability evaluation methods usually cover the following common characteristics (Shneiderman and Plaisant, 2004):

- ease of learning
- speed of user task performance
- rate of errors by users
- subjective user satisfaction
- user retention of commands over time

Nielsen (1994a), generalising from usability evaluations of eleven software applications from various fields and approximately two hundred and fifty usability problems, produced ten heuristics for usable design:

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognise, diagnose, and recover from errors
- Help and documentation

Despite the apparent extensiveness of these heuristics, simply applying evaluation methods derived from usability research studies to the study (and evaluation) of eLearning applications will not meet the needs and purposes of the TELT Evaluation Framework. Therefore, a more extensive correlation of the results of studies that match Homs’ (1998) three categories of evaluation is required.

While the preceding heuristics are suitable for a general evaluation of all types of eLearning application software, there are alternative methodologies designed for specific categories of eLearning applications, many of which are based on heuristic design techniques. The remainder of Section 1.1 describes the evaluation methodologies identified in this study as applied to the following categories of eLearning applications:

- learning management systems (LMS)
- synchronous Computer-mediated Communication systems (CMC)
- plagiarism detection software
- computer assisted assessment tools
- educational multimedia software

### 1.1.2 Learning Management Systems (LMS)

Methodologies for evaluation of the LMS reviewed for this section cover the typical life cycle of an institution’s usage of a LMS, namely: LMS selection; educational design on LMS; learning and teaching activities; and, users’ overall satisfaction of their LMS experiences. The methodologies revealed in the literature that fit into these phases have different foci and use varying approaches to achieve the evaluation objectives.

Obexer (2005) and the Curriculum Corporation and Educationan.au Limited (2002) focused on the ideal functions of a LMS whereas Britain and Liber (2004) evaluated LMS use from a pedagogic perspective. The evaluation processes structured by these methodologies can be implemented without actual end users’ participation. Mehlenbacher et al., (2005) draws attention to the educational design implemented on a LMS that may reflect the extent to which it assists the educational / instructional design process. Newland, Norman, Hall, Wensley, and Wardle (2004) evaluated learning and teaching activities that were supported by a LMS in Bournemouth University. Their evaluation study covered educational value and the usefulness of the LMS. Finally, Wang (2003) measured users’ satisfaction on their experience with a LMS. Table 1 provides an overview of the studies reviewed to his point.

**Table 1: Overview of the Methodologies Reviewed in Section 1.1.2**

<table>
<thead>
<tr>
<th>Document Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework Stage 3, sub-layer 1</th>
</tr>
</thead>
</table>

42
1. Obexer (2005) | Expert review, reference check, hand-on evaluation | Checklist | Could be modified into a user survey for measuring users’ perception towards the functions listed


3. Mehlenbacher et al., (2005) | Heuristic evaluation | Usability heuristic questions | Could be used for evaluating eLearning design for courses that participate in pilots. The concern is that the eLearning designer’s influence may dominate the LMS per se and therefore, the result may not necessarily reflect the application’s educational value and usefulness.

4. Britain & Liber (2004) | Heuristic evaluation | Questionnaire and semi-structured interviews | The criteria and associated heuristic questions could be readily adapted to questionnaires and interview questions for end-users to answer.


6. Wang (2003) | Survey | Questionnaire | Could be used after the implementation of pilots for the users to complete.

7. Ho, Higson, Dey, Xu, & Bahssoon (2009) | Integrated multiple-criteria, decision making approach | Accumulate criteria from stakeholders | This methodology is to collect the evaluation “criteria” from stakeholders so that the most appropriate LMS can be selected. This methodology can be used for other types of eLearning applications.

1. The Queensland University of Technology (Obexer, 2005) developed a list of functional evaluation criteria for learning and teaching, which formed part of their LMS evaluation criteria and strategies. The criteria were divided into eleven categories:

- Content creation and Management
- Communication and Collaboration
- Learning design and Pathways
- Assessment and Grade Management
- Student tools
- Navigation and Interface
- Accessibility
- Copyright
- Tracking and Analysing
- Permissions and Security
Innovation and New technologies

The eleven categories were specifically designed for evaluating learning management systems and have an apparent focus on the applications’ features. For the most part, the result of the evaluation study may identify the comprehensiveness of the range of available functions, but rarely addresses their educational implications. One criterion under the Communication and Collaboration section is Wikis: the learning management system is expected to have a wiki system for students’ collaboration. However, the inclusion of a wiki tool does not guarantee ease of use and capacity to meet educational needs. Taking Moodle as an example, one of the concerns about its wiki system is the difficulty for students to insert an image into a wiki page (Okada, 2008).

2. The evaluation framework devised by the Curriculum Corporation and Education.au Limited (2002) represents another function-focused methodology for evaluating learning management systems, which is similar to Obexer (2005). The instrument is a set of Likert four-point scale evaluation criteria that were organised through three layers:

- Business layer: the vision, objectives, drivers, key functions, processes, organisations and their relationships.
- Systems layer: the information and applications that support the business level.
- Technology layer: elements such as hardware and networks, integration and security facilities that support the systems level.

The names of the layers do not share the same meanings of the terms used in the TELT Evaluation Framework. The so-called Business and Systems layers only provide broader and more general description of relevant criteria in the Technology layer. So the criteria that are actually considered in the evaluation are suited to the Technology layer, which contains a range of ideal functions and services that learning management system packages should have. The following are the top-level criteria:

- Flexibility, interoperability, reusability, efficiency and security
- Expertise, support and maintenance
- Access management, student management, administration and tracking within learning systems
- Whole-of-school ICT management: flexibility, interoperability, reusability, efficiency and security
- Teaching and learning: teachers, students and their families
- Fostering collaboration and communication.

3. Mehlenbacher et al., (2005) developed a conceptual model for designing and evaluating eLearning programs developed in a learning management system, which informs five dimensions that are present in all instructional contexts:

- Learner Background and Knowledge
- Learner Tasks and Activities
- Social Dynamics
- Instructor Activities
- Learning Environment and Tools.
The model advanced by Mehlenbacher et al., (2005) supplies a list of usability heuristic questions grouped into the above five dimensions. The main feature of this instrument is that the heuristics reflect instructional activities. The following are sample questions that illustrate this point:

- Do menu instructions, prompts, and error messages appear in the same place on each screen?
- Are learners rewarded for using the communication tools?
- Are levels clear and explicit about the “end” or parameters of the site?
- Is information structured by meaningful labelling, bulleted lists, or iconic markers?
- Is an overall tone established that is present, active, and engaging?
- Does the site use an easily recognisable metaphor that helps users identify tools in relation to each other, their state in the system, and the options available to them?

4. Based on the Conversational Framework (Laurillard, 1993) and the Viable System Model (Beer, 1979, 1981, 1985), Britain and Liber (2004) developed a pedagogical framework to evaluate learning management systems. Evaluation conducted within the framework covers all activities in the teacher-medium-student conversation cycle and emphasises three perspectives:

- Managing teaching and learning in a module or course (Module Level)
- Student management of their own learning (Learner Level)
- Management of modules within an overall course or degree programme at an institutional level (Institution Level)

The framework uses the above three levels to organise a list of criteria and associated heuristic questions. The methodology has been tested by a series of case studies in which the WebCT Vista 2.1, Blackboard Academic Suite 6.1, Granada Learnwise 3, and FirstClass 7.1 were evaluated through semi-structured interview based on the questions derived from the framework. The interviewees were the vendors or creators of the learning management systems. The following criteria were used in the evaluation:

- Module
  - presentation and re-presentation of key concepts and ideas
  - coordination of people, resources and activities
  - resource negotiation and agreement
  - monitoring of learning
  - self-organisation amongst learners
  - adaptability of module and system
- Learner
  - learner-centeredness
  - coordination of people, resources and activities
  - time management / planning
  - monitoring own learning
  - adaptation / reflection
- Programme / Institution
  - extensibility and integration
  - coordination of people and activities
  - resource negotiation and agreement
  - monitoring of modules
To adapt Britain and Liber’s (2004) evaluation framework to Stage 3, sub-layer 1 of the TELT Evaluation Framework, the criteria and associated heuristic questions can be redeveloped into questionnaires and interview questions as appropriate.

5. Newland et al., (2004) reported on a study of the Bournemouth University (UK) learning management system that covered both educational value and the usefulness of the application. Data were collected from students and teachers, analysed separately, and then compared to reveal a complete picture of learning and teaching delivered through the learning management system. The methods applied to the evaluation study involved automatic data collection and questionnaires. The criteria used in the questionnaires were as follows:

- **Students**
  - **Access and Usage of LMS:**
    - frequency of use
    - location of access
    - confidence using LMS
    - help documentation
    - preference for materials
    - printing and reading behaviours
    - recommended use of LMS by lecturers
    - use of media2 by learning support material
  - **Effects on Learning:**
    - ability to see material on other programme pages
    - learning effect by learning support material
    - note-taking behaviour
    - benefits of LMS within lecture framework
    - additional benefits of LMS
    - overall contribution LMS has made to learning

- **Teachers**
  - **Access and Usage of LMS:**
    - frequency of use
    - motivation for using LMS
    - units used to support teaching
    - confidence using LMS
    - ease of uploading material
    - training
    - maintaining/updating
    - amount of content
    - resources made available to students
  - **Effects on Teaching:**
    - note-taking behaviour
    - quality of teaching
    - online participation
    - approach to teaching
    - clarity of student learning pathway
    - ability to see other material
management, collaboration and communication
overall effectiveness of LMS

Bournemouth University’s experience in terms of the structure of the evaluation provides useful direction for the TELT Evaluation Framework proposal, which in part has been taken into account.

6. Wang’s (2003) model for evaluating learning management system focuses on learner satisfaction. The model has twenty six criteria that focus on the interactions between the system and users, which is divided into four groups:

- learning interface
- learning community
- content
- personalisation

The model's reliability and validity have been assessed against (1) the intended reuse of the learning management system and (2), the extent of post-usage compliance to the stated reuse. The primary benefit of the model is that it provides a ready-to-use instrument for evaluating user satisfaction.

7. Ho, Higson, Dey, Xu, and Bahsoon (2009) developed an integrated multiple criteria decision-making approach to evaluate and select learning management systems. The proposed methodology combines a hierarchical analytic process (AHP) and quality function deployment (QFD) as described in the procedural steps suggested by Ho et al., (2009):

- Identify the stakeholders who use the LMS
- Determine the proportion of each stakeholder category
- Identify the stakeholder requirements
- Determine the relationship weightings between the stakeholders and the stakeholder requirements using AHP
  - AHP pair-wise comparison
  - AHP synthesisation
  - Compute the average of the entries in each row of matrix A’ to yield column vector
  - AHP consistency verification
  - Compute the averages of values in vector ĉ to yeild the maximum eigen value of matrix A
  - Compute the consistency index
  - Compute the consistency ratio
- Compute the importance rating of each stakeholder requirement
- Copy the stakeholder requirement and their corresponding importance ratings into HOQ2
- Identify the LMS features
- Determine the relationship between weightings between LMS features i and its corresponding stakeholder requirements k, c_i^k using AHP
- Compute the importance rating of each LMS feature
- Copy the LMS features and their corresponding importance ratings into HOQ3
- Identify the LMS alternatives
- Determine the relationship weightings between LMS i and its corresponding LSM features k, c_i^k using AHP
Calculate the total score of each alternative LMS:

\[ w_i^2 = \sum_{k \in S} \bar{w}_i^2 c_{ik} \]

The feature of Ho et al.'s. (2009) methodology that distinguishes it from other methodologies is that they did not pre-define the criteria for determining a more effective LMS. Instead, their strategy was to collect the “criteria” from stakeholders so that the most appropriate LMS can be selected from the results of evaluation process. This methodology can also be used in other types of eLearning applications. The drawback is that it requires complex quantitative computing.

1.1.3 Virtual Meeting Rooms and Virtual Campuses (Synchronous CMC Systems)

Virtual meeting rooms and virtual campuses are online environments where users interact and collaborate synchronously with each other from their own computer terminals using audio, video and text-based chat and application sharing tools. Such tools are also referred to as Computer-mediated Communication systems (CMC). This type of interaction can be integrated with learning management system that either enable learning management features in virtual environments or enrich the virtual meeting room functions in the learning management system. All the reviewed methodologies for evaluation of synchronous CMC systems involve the use of questionnaires to enquire what users think of the tools, but the results reported in most publications vary. Islam (2009) focuses on user satisfaction; De Lucia et al., (2009) measured the CMC outcomes; Chou (2001) was concerned about users’ perception towards various functions of different CMC tools; Ras, Angkasaputra, and Waterson (2004) used a user group comparison approach to identify the benefits that a virtual meeting room brought forward to research projects. Table 2 provides an overview of the documents reviewed in this section.

Table 2: Overview of the Methodologies Reviewed in Section 1.1.3

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Islam (2009)</strong></td>
<td>User review</td>
<td>Questionnaire</td>
<td>To utilise the instrument, it has to be significantly modified by adding more criteria for issues that emerge when using virtual meeting tools.</td>
</tr>
<tr>
<td><strong>2. De Lucia et al., (2009)</strong></td>
<td>User self-reflection &amp; user review</td>
<td>Questionnaires</td>
<td>The questionnaires are suitable for students to complete who participate in pilots, so that the effectiveness of how well the virtual environment supports learning can be measured.</td>
</tr>
</tbody>
</table>
### 3. Chou (2001)

<table>
<thead>
<tr>
<th>Expert observation, user self-evaluation, and user review</th>
<th>Questionnaires</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are obvious weaknesses in terms of evaluation criteria and overlapped items in questionnaires. The questionnaires could be modified to be used for monitoring students’ confidence in specific tasks when participating in CMC sessions over a period of time.</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Experimental and control groups comparison</th>
<th>Questionnaire and sub-task monitoring list</th>
</tr>
</thead>
<tbody>
<tr>
<td>The method for measuring effort and duration can be further developed for comparison of different virtual meeting rooms while TELT is piloting more than one virtual meeting room application.</td>
<td></td>
</tr>
</tbody>
</table>

1. Based on Doll and Torkzadeh’s (1988) information system user satisfaction model, Islam (2009) developed a new model to measure user satisfaction and the success of virtual meeting tools. The model is a five-point Likert questionnaire that covers the following areas of measurement:

- Content
- Accuracy
- Format
- Ease of use
- Timeliness
- Reliability
- Speed

Despite the fact that the reliability and validity of Islam’s (2009) model have been tested based on empirical studies where internal and external participants used the virtual meeting tool and completed the questionnaire, this instrument still seems less than mature. Its foundation was a general information system satisfaction measurement and only added two more dimensions to Doll and Torkzadeh’s (1988) model (reliability and speed). The evidence that the model was tailored to specifically evaluate virtual meeting tools is not obvious. So when adapting it, more criteria should be added to this model, which should focus on the issues that emerge in the use of virtual meeting tools.

2. De Lucia et al., (2009) evaluated the pilot lectures delivered through a virtual campus built in Second Life. The procedure involved pre- and post-questionnaires completed by the students, which were adapted from Witmer and Singer (1998). The structure of the questionnaires is as follows:

- User background (pre-experiment questionnaire)
  - PC knowledge
  - 3D environments and PC games knowledge
  - Tendency to become involved in activities
- Presence
  - Control factors
  - Sensory factors
  - Distraction factors
  - Realism factor
  - Involvement
  - Haptic

---

49
• Communication
• Awareness
• Perceived sociability
• Virtual environment
• Productivity and general satisfaction

De Lucia et al.'s, (2009) evaluation was based on the assumption that in a 3D multi-user virtual environment, learning is strongly related to the user's sense of belonging to a learning community, as well as their perceived degree of awareness, presence and communication. The results of the evaluation were very positive. The design of this methodology reflects the link between the support of the technology and the factors that lead to effective learning and is therefore suitable for being adapted to the TELT evaluation framework.

3. Chou (2001) conducted an evaluation of a range of synchronous CMC systems used in a course offered by the University of Hawaii in USA, which included three types of tools: (1) text-based conferencing systems: ICQ and WebCT chat; (2) audio-video conferencing systems: Netscape CoolTalk and CU-SeeMe, and (3), enhanced virtual systems:

ActiveWorlds (http://www.activeworlds.com) and The Palace (http://www.thepalace.com).

The evaluation methodology was based on learner-centred principles, constructivist and socio-cultural theories. The following evaluation methods were used:

• Expert observation: four observers kept a log of their observations of all the instructional activities in the class, focusing on three areas:
  • student development
  • instructor-student interaction
  • course elements
• Ratings on CMC skills: students filled out a self-evaluative rating sheet on their CMC skills at the beginning, middle, and end of semester on a one-to-five Likert scale for the following items:
  • E-mail
  • Newsgroups
  • Listservs
  • CU-SeeMe
  • MOO/MUD
  • Electronic meeting room
  • Web search
  • Web-based courseware
  • Web chat
  • Internet-based audio-conferencing
  • Internet-based video-conferencing
  • HTML editor
• Ratings on CMC systems: after each use of the CMC system, students were asked to complete an online rating sheet on three areas:
  • Social presence
• Communication effectiveness at an interpersonal level and at a system level
• Communication interface

Based on the findings and discussions, Chou (2001) developed a checklist of critical features of an effective CMC system:

• Low bandwidth
• Ease of navigation
• Accessibility
• Non-intrusiveness
• Affective affirmation
• Fun and pleasure
• Humanising and sensitive qualities
• Good audio-video quality
• Support tools for knowledge construction
• Community building

Chou (2001) applied multiple approaches to evaluate a range of CMC systems from various aspects. The evaluation approaches and procedures provide a useful example for application to the TELT pilot projects. However, the obvious limitation of Chou’s work is the lack of generalised evaluation criteria for all types CMC systems and specification of the differences in evaluation strategies among the three types of CMC tools examined. Another weakness is the overlap of assessment heuristics for monitoring students’ progress in CMC skills. However the rating sheet devised by Chou could be further developed to monitor students’ confidence in carrying out specific tasks when participating in CMC sessions over a period of time. The result may indicate the ease of use of the tested systems.

4. Ras et al., (2004) evaluated the virtual meeting room built in a collaborative learning system. Participants were divided into two groups: a supported group that used the virtual meeting room to develop a research profile and an unsupported group that only used emails and face-to-face meetings for another task with the same level of complexity. The supported group was asked to complete the rating of the tool against the following heuristics:

• The tool enabled me to perform my tasks.
• The use of the tool increased the quality of my contribution.
• The navigation through the conversation was easy.
• The tool provided useful support to my awareness about other people.
• I found it easy to communicate with the other participants.
• I found it easy to use the tool.
• The user interface was intuitive for me.

Both groups reported on their effort (person hours) and duration (working days) put into various sub-tasks and then Ras et al., (2004) compared the total effort and duration. The sub-tasks that were monitored included:

• Supported group
  • Preparation of meeting
  • Carrying out the face-to-face meeting
- Carrying out individual interviews between moderator and the group members to finish the application profile table
- Unsupported group
- Conducting a tool tutorial including self-practice with the virtual meeting room
- Preparation for the meeting
- Carrying out the meeting using the virtual meeting room

Ras et al.'s., (2004) method for measuring total effort and duration can be developed for comparison of different virtual meeting room as these two criteria may indicate the efficacy of CMC tools for group learning.

### 1.1.4 Plagiarism Detection Software

Plagiarism detection software assists lecturers to detect plagiarism evident in students’ submitted work by comparing the work against online resources, publication databases, and/or peers’ work. To determine whether a plagiarism detection tool is successful, three main aspects should be evaluated—(1) the plagiarism detection quality (e.g., high plagiarism rate should be reported for a seriously plagiarised work, and vice versa), (2) the operation of the software and (3) users perceptions of the software. The reviewed methodologies indicate that the main approach to measure the quality of plagiarism detection software is testing.

For the other two aspects, researchers normally use a qualitative approach to collect data from users and then generalise the analysed results. However, the reviewed literature does not cover the situation where students are given the right to pre-check their work. This type of service is actually available (for example, [http://writecheck.turnitin.com](http://writecheck.turnitin.com)). Instead of detecting plagiarism, the pre-check service helps students to cite their sources appropriately, which means that the focus is on educational function rather than regulation, monitoring or punishment. This strategy is worthy of consideration for the TELT platform.

Table 3 outlines the documents reviewed in this section.

### Table 3: Overview of the Methodologies Reviewed in Section 1.1.4

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hashem (2006)</td>
<td>Testing and comparison</td>
<td>Evaluation criteria</td>
<td>May be of use for developing testing criteria. Hashem (2006) also provided us with a list of weakness of the testing method, which we should be aware of.</td>
</tr>
<tr>
<td>2. Arts &amp; Geus (2004)</td>
<td>Testing and expert review</td>
<td>Evaluation criteria and evaluation form</td>
<td>The evaluation form for each application is comprehensive and covers most of the important dimensions, which could be directly used for developing the TELT Platform Evaluation Framework.</td>
</tr>
<tr>
<td>3. Bull, Collins, Coughlin, and Sharp (2001)</td>
<td>Testing and user review</td>
<td>Evaluation criteria and evaluation form</td>
<td>The evaluation criteria cover most of the areas that directly reflect the quality of the software. They are of genuine use for TELT.</td>
</tr>
</tbody>
</table>
4. **Chester (2001)**

Six-month pilots in multiple institutions followed by user review

N/A

The structure of the evaluation report is useful for developing report templates. The findings from the unstructured pilot evaluation are valuable and can be further developed for use as criteria for structured evaluation.

5. **California State University at Chico (2004)**

User survey

Questionnaire

The questionnaires were designed according to two main themes—frequency and satisfaction of usage of the software. These two themes could be integrated into other criteria for the evaluation of plagiarism detection software.

6. **Scaife (2007)**

Vendor self-evaluation, testing and expert review

Vendor self-evaluation questionnaire

The vendor self-evaluation questionnaire could be modified to provide a list of evaluation heuristics. The testing method is of little use as there is no common ground to compare the plagiarism scores reported by each tool.

1. Hashem (2006) conducted a research study to compare two plagiarism detection software products: MyDropBox and Pl@giarism to which the same essays were submitted and the detection reports were compared against the following criteria:

   - Number of essays using freely available material from the Internet without proper citation
   - Number of pairs of essays containing similar material, with perhaps minor changes
   - Number of essays using material from proprietary databases and subscription-based online libraries.

   Hashem (2006) also highlighted the weakness of this method:

   - The three criteria may overlap. For example, one essay may contain material from another essay, which in turn contains material from the Internet.
   - Computer programming is not supported by the software. Other plagiarism detection software specialised in computing programming should be used.
   - Materials from printed books are not generally supported for detection.
   - Translations of materials into other languages are not detectable.

2. Arts and Geus (2004) compared the performance of six plagiarism software solutions: Edutie, Plagiserve, TurnItIn, EVE2, Wcopyfind, and Pl@giarism. 10 randomly selected documents were processed through the tools and evaluated against the following criteria:

   - Quality (of the results/reports)
   - Stability of the system
   - Speed of the response
   - General user friendliness and ease of use.

   The evaluation conducted by Arts and Geus (2004) mainly used a qualitative approach to produce a profile for each software, which includes the software’s name, URL, detection type, vendor, demo/trial version, price, platform, submission format, scope of
3. Bull, Collins, Coughlin, and Sharp (2001) used a combination of qualitative and quantitative means across various institutions to evaluate Findsame, Eve2, Turnitin, CopyCatch and WordCHECK from two perspectives:

- Academic user perspective
- Function
- Detection performance (The scores awarded reflect the performance of the tools to detect the known instances of plagiarism in accordance with the performance as stated in the products' literature and website)
- Clarity of reports
- Overall feel/user friendliness
- Detection at level
- Obtaining the tool
- Clarity of instruction and support materials
- Submitting documents
- Can results be printed?
- Accuracy of detection
- Layout of reports
- Are results easy to interpret?
- Accuracy of results
- Technical perspective
- Operating environment
- Ease of mass distribution
- Turnaround speed
- Installation engine
- Reliability
- Suitability for mass distribution
- Stability of vendor
- Speed of response
- Technical support

4. Chester (2001) qualitatively evaluated the pilot of iParadigms in five participating institutions across Europe. Five subjects on offer from each institution used iParadigms to detect plagiarism over a six-month period. Instead of focusing on the software per se, Chester (2001) was concerned with the broader issues and impacts, for example, the impact of training both staff and students on plagiarism issues, the impact on institutions’ current plagiarism policies and procedures, interactions with student groups, possible cultural issues, possible appeal strategies, possible issues surrounding the isolated nature of distance learning. The main methods used in this evaluation were that each institution was required to provide a report detailing the experiences of using the software, and the project team conducted workshops to discuss findings from each institution. The following are the themes of the findings:

- Department withdrawing from the pilot
- Service accessibility
- Reaction of student to the project
• Reaction of staff to the project
• Software strengths
  • Growing Database of Material
  • Ease of Use
  • Student Use of the Web
• Software weakness
  • Technical response
  • Login process
  • IT Skills required
  • Speed of result
  • Upload of text
  • Browser compatibility
  • Lack of detection of published materials (books, journals, etc.)
  • Inability to detect images
• Legal issues
  • Purchasing a service overseas
  • Students’ copyright permission
• Central versus local detection service
• Issues to consider when selecting detection software
  • Ease of Use
  • Cross-referencing and collusion identification
  • Uploading of Material
  • Browser specifications
  • Re-analysis Facility
  • Response time of software supplier
  • Integration with Existing Systems

Chester (2001) generalised the experiences produced from the unstructured pilot evaluation. The findings are valuable and are suitable for further development into criteria for structuring an evaluation model.

5. California State University at Chico (2004) conducted a structured web-based survey to collect data on academic staff experiences in using TurnItIn. The survey applied frequency and satisfaction of usage scales to determine the software’s usefulness for learning and teaching. The following criteria are reflected on the survey:

• Frequency of assigning graded student activities where plagiarism could be an issue
• Frequency of and satisfaction with methods to detect plagiarism:
  • using TurnItIn.com
  • using Plagiarism detection services other than Turnitin.com (e.g., CopyCatch, EVE2)
  • using Free WWW search engines (e.g., Google, Yahoo, AltaVista) to detect plagiarism
  • using proprietary databases provided by the university library to detect plagiarism
  • using own books, periodicals, or other materials to detect plagiarism
  • using library books, periodicals, or other materials to detect plagiarism
  • using assistance from the librarians in the university library to detect plagiarism
  • using other methods to detect plagiarism
• Satisfaction with aspects of TurnItIn:
  • As a method of detecting plagiarism
  • As a method of preventing/discouraging plagiarism
  • Intuitiveness of the Turnitin.com interface
6. Scaife (2007) conducted a comprehensive evaluation of ten plagiarism detection applications, which were divided into four classifications:

- Check against Internet based material only
  - DOC Cop
  - Docoloc
  - EVE2
  - Plagiarism Checker
  - Scriptum / CatchItFirst
- Check for similarity within a batch of documents
  - DOC Cop
  - WCopyfind
- Educational suite of applications offering complete student submission solutions
  - Ephorus Suite
  - MyDropBox
  - Turnitin
  - Urkund
- Other
  - Glatt

The methods involved in Scaife's (2007) evaluation were a questionnaire and testing. The questionnaire was completed by the vendors. It consisted of eighty six close-ended and open-ended questions organised into the following sections:

- Installation / Configuration
- Stability of application
- Databases searched
- Usability
- Upload options
- Output options
Scaife (2007) awarded a score for each section of the questionnaire and it turned out that TurnItIn was awarded the highest score. However, in the report, Scaife (2007) did not articulate how the scores were allocated based on the vendors’ responses to the questions.

As for the testing process, Scaife (2007) used each plagiarism tool to test different files with different expected plagiarism scores to determine whether the reported scores matched the initial expectations. The problem encountered during the testing phase was that there was no common ground to compare the plagiarism scores reported by each tool. For example, an 80% score from one product could be equivalent to 100% with another.

1.1.5 Computer Assisted Assessment (CAA) Tools

Computer Assisted Assessment tools refer to software solutions that assist lecturers to develop online / print-based assessment materials, mark students’ work and analyse assessment results. Based on the documents reviewed, formal evaluation approaches for CAA tools are usually heuristics-based, but the emphases of the heuristics vary. Jones (2007) focuses on functions supporting e-marking and feedback; Sim et al. (2009) is all about usability; and Sim et al. (2007) measured user satisfaction. Most of the heuristics described in the literature are of genuine value for the TELT Evaluation Framework. Table 4 provides an overview of the documents reviewed in this section.

Table 4: Overview of the Methodologies Reviewed in Section 1.1.5

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework Stage 3, sub-layer 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. University of Leeds (n.d.)</td>
<td>Function comparison, comparison table</td>
<td>Comparison table</td>
<td>Could be further modified into an instrument that matches academics’ assessment needs and the usability of assessment tools.</td>
</tr>
<tr>
<td>2. Jones (2007)</td>
<td>Heuristic evaluation and survey, heuristics and questionnaire</td>
<td>Heuristics and questionnaire</td>
<td>From the evaluation result, Jones (2007) generalised 4 themes to be considered when making a decision to purchase e-marking tools. These 4 themes along with the heuristics could be further developed into evaluation criteria.</td>
</tr>
<tr>
<td>3. Sim et al. (2009)</td>
<td>Heuristic evaluation, heuristics</td>
<td>Heuristics</td>
<td>The heuristics employed by Sim et al. (2009) are highly useful and readily valuable for application to the TELT Evaluation Framework.</td>
</tr>
</tbody>
</table>
The University of Leeds in UK (n.d.) developed a checklist to compare two assessment tools—Questionmark Perception and Blackboard. The checklist contains two sections: (1) question types supported by the tools and (2) other features such as adding text in between questions and adaptive branching. On the checklist, academic staff are advised to choose one tool that suits their needs, but the checklist does not indicate any judgement of the tools. A procedure or a mechanism of matching academics’ assessment needs and the usability of assessment tools needs to be developed.

Jones (2007) focused on the marking and feedback processes of assessment assisted by technology, and evaluated four e-marking tools for use in Higher Education from the perspectives of staff and students. The tools evaluated were Electronic Feedback, M2AGICTM, GradeMark, and e-Rubric. The criteria used for the evaluation included:

- Does the tool support marking criteria?
- Does the tool support marking standards?
- Is it clear to the student in the output generated by the tool how marks have been awarded?
- Does the tool enable feedback for learning?
- Does the tool enable the provision of feedback that is positive?
- Can the tool be used to provide feedback and marks for different types of assessment?
- Can the tool be used to mark and provide feedback for group work?
- Can the tool be used to provide self and/or peer assessment?
- Does the tool promote efficiency of marking?
- Does the tool enable marking to be managed / monitored by the module coordinator?
- Does it provide feedback that allows tutors to shape their pedagogic practice?
- Can the tool be used for anonymous marking?
- Can the tool be used for marking hard copy? And, can the tool be used for electronic submissions?
- Is the tool easy to use?
- Is the tool free to use?
- What are the computer or technical requirements?
- Is the tool web based?

The justification for these criteria was indicated to be available in the full report on the project, but it is not available for download. The methods involved in the evaluation included focus group interview and questionnaires throughout the pre- and post-testing processes. Jones (2007) compared the four tools, highlighted the issues associated with each tool and generalised five themes to be considered when making a decision to purchase e-marking tools. The themes are:
• The need for staff to be computer literate
• Familiarisation and preparation time
• Attaching marks to criteria
• Feedback to students
• Flexible use,

3. Sim, Read, and Cockton (2009) examined the effectiveness of Nielsen’s (1994b) heuristics set in the context of computer assisted assessment, built the corpus of usability problems by testing three CAA applications (Questionmark, WebCT and TRIADS), and then developed a revised and extended CAA-specific heuristic set for evaluating computer assisted assessment tools. The following are the outcomes:

• Same as Nielsen’s Heuristics
  • Error prevention and recovery
  • User control and freedom
  • Ensure appropriate help and feedback
• Modification of Nielsen’s Heuristics
  • Ensure appropriate interface design characteristics
  • Answering question should be intuitive
• New Heuristics
  • Navigating within the application and terminating the exam should be intuitive
  • Prevent loss of input data
  • Accessing the test should be clear and intuitive
  • Use clear language and grammar within questions and ensure the score is clearly displayed
  • Design should inspire trust and not unfairly penalise
  • Minimise external factors which could affect the user

Additional descriptions for each of the above heuristics are available in Sim et al.’s (2009) study. The heuristics set was developed based on the widely recognised Nielsen’s (1994b) heuristics set and empirical data from over ninety evaluators and three hundred students. It is a ready-for-use instrument for evaluating CAA applications and is of immediate use for the TELT Evaluation Framework.

4. Sim, Read, and Holifield (2007) conducted an evaluation to measure the user satisfaction of three CAA applications. Each of the undergraduate students who participated in the study completed three tests using three different CAA software environments. User satisfaction was measured through the use of two survey instruments, which were based on previous studies. The questions in the survey relate to five types of user experiences:

• Accessing and Finishing the Test
• Visual Layout
• Navigation
• Answering the Questions
• Preference for Software Depending on Context
The two evaluation instruments devised by Sim et al. (2007) cover most of the areas of student users’ experiences of CAA applications. However, CAA application users are not limited to students. Teaching staff and administrative staff should also be taken into account. Therefore, to utilise these instruments for the TELT Evaluation Framework to measure user satisfaction comprehensively, the surveys require further modification to be useful for the TELT platform.

5. Khedo (2005) summarised the benefits of CAA from administrative and pedagogical perspectives, which may be useful for developing surveys to investigate staff and student users’ perception of the benefits that CAA pilots afford:

- **Administrative benefits:**
  - Large groups are more easily assessed.
  - Marking is not prone to human error, marker fatigue or subjectivity.
  - Computerised marking of tests can result in a great saving of staff time in the longer term.
  - There is a reduction in printing costs.
  - Simplified administration, once test formats and question banks are set up.
  - It can aid in the transmission and administration of marks that can be automatically entered into information management systems and student records databases.

- **Pedagogic benefits:**
  - Students can monitor their own progress; self assessment can be promoted.
  - Adaptive testing can be used to match test to the student’s ability.
  - Diagnostic reports and analyses can be generated—for example, to see which areas of the curriculum are not understood.
  - Quicker, more detailed, and more accurate feedback can be given to both student and lecturer.
  - Increased scope for statistical analysis of test outcomes. The results are stored in a format that is amenable to the rapid production of statistical information needed by lecturers to analyse students’ performance in a test.
  - Lecturers can monitor the progress of students through more frequent assessments.
  - A wide range of topics within a body of knowledge can be tested very quickly.
  - The potential to introduce graphics and multimedia allows for inclusion of questions not possible with paper assessments.
  - Cheating can be reduced through the randomisation of questions.
  - Students report reduced stress levels when sitting for computer-assisted exams as compared to traditional examination venues.
  - It can provide lecturers with feedback for evaluation of modules or courses.

Khedo (2005) also summarised the technical issues associated with the implementation of CAA applications. This will be discussed in the Technical Layer section of this paper.

1.1.6 Educational Multimedia Software

Educational Multimedia software use a combination of different content forms (text, audio, still images, animation, video, and interactivity) to deliver instruction. Many studies about multimedia in education emerged in the 1990s, among which evaluation was one of the popular topics. The main methodology for the evaluation of multimedia was heuristic evaluation (Crozat et al. 1999, Park and Hannafin 1993, Albion 1999, Pham 1998, and
Heller, Martin, Haneef, and Gievska-Krliu (2001). The heuristics used in these studies covered a wide range of areas that normally should be the focus of evaluation studies. However, it possible to identify a number of overlaps and therefore it would be useful to integrate all of these heuristics (where possible) as the basis for devising suitable heuristic for the TELT evaluation framework.

Other researchers in this area such as Barker and King (1993) used a survey approach to conduct an evaluation that used two versions of questionnaires for different groups of users. Kennedy, Petrovic, and Keppell (1998) and Heller et al., (2001) provided sound theoretical frameworks for developing other types of evaluation instruments. Table 5 outlines the documents reviewed in this section.

Table 5: Overview of the Methodologies Reviewed in Section 1.1.6

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crozat et al. (1999)</td>
<td>Heuristic evaluation</td>
<td>Heuristics</td>
<td>The EMPI model is suitable for evaluating multimedia eLearning software where there are built-in instructional materials.</td>
</tr>
<tr>
<td>2. Park &amp; Hannafin (1993)</td>
<td>Heuristic evaluation</td>
<td>Heuristics</td>
<td>The 20 educational principles and their implications for designing and evaluating interactive multimedia for education can be further developed to be a heuristic evaluation instrument that assists in determining to what extent the multimedia applications are meeting the stated educational principles.</td>
</tr>
<tr>
<td>3. Barker &amp; King (1993)</td>
<td>Survey</td>
<td>Questionnaires</td>
<td>The two versions of questionnaires (one for experts and one for novice users) are suitable for self-evaluation as well as TELT pilots. They are ready for use.</td>
</tr>
<tr>
<td>4. Albion (1999)</td>
<td>Heuristic evaluation</td>
<td>Heuristics</td>
<td>The heuristic list covers the three important areas of educational multimedia development — interface design, educational design, and content design. It is ready for use.</td>
</tr>
<tr>
<td>5. Pham (1998)</td>
<td>Heuristic evaluation</td>
<td>Heuristics</td>
<td>The questions that Pham (1998) raised may be treated as a general reference only because Pham’s (1998) thoughts have not formed a clear framework, nor did Pham (1998) actually use the proposed questions to evaluate multimedia applications. The questions were led by general discussion about different aspects of the quality of multimedia systems.</td>
</tr>
<tr>
<td>6. Kennedy et al., (1998)</td>
<td>N/A</td>
<td>Theoretical framework</td>
<td>Kennedy et al., (1998) did not supply exact evaluation criteria. Instead, they justified the use of four evaluation areas that were supported by research literature. The terms, scopes and meaning of the four areas articulated by Kennedy et al., (1998) would be useful for developing the TELT...</td>
</tr>
</tbody>
</table>
criteria framework.

| 7. Heller et al., (2001) | Heuristic evaluation | Multimedia taxonomy for evaluation | This framework is of great use for us to develop evaluation criteria for multimedia applications on the TELT platform. |

1. Crozat, Hu, and Trigano (1999) developed a multimedia evaluation model called EMPI (Evaluation of Multimedia, Pedagogical and Interactive software). EMPI consists of a list of evaluation criteria and sub-criteria grouped into six modules along with examples of associated questions:

   - **General feeling**: what image the software offers to the users
   - **Technical quality**: technical realisation of the software
   - **Usability**: corresponds to the ergonomics of the interface
   - **Multimedia documents**: text, sound, image, etc.
   - **Scenario**: the writing techniques used in order to design information
   - **Didactics**: pedagogical strategy, tutoring, situation, etc.

   The heuristic questions in the EMPI model are suitable for evaluating multimedia eLearning software where there are built-in instructional materials. However, some of the heuristics would only result in a neutral descriptive statement rather than a judgement. The evaluator has to decide if the neutral responses are positive or negative in the particular context.

2. Park and Hannafin (1993) reviewed empirical studies on interactive multimedia with psychological, pedagogical and technological foundations, and then summarised twenty educational principles and their implications for designing and evaluating interactive multimedia for education. The following are the implications:

   - Layer information to accommodate multiple levels of complexity and accommodate differences in related prior knowledge.
   - Embed structural aids to facilitate selection, organisation, and integration; embed activities that prompt learners to generate their own unique meaning.
   - Organise lesson segments into internally consistent concepts
   - Linkages between and among nodes need to reflect the diverse ways in which the system will be used
   - Provide opportunities to reflect critically on learning and to elaborate knowledge; encourage learners to articulate strategies prior to, during, and subsequent to interacting with the environment.
   - Use familiar metaphors both in conveying lesson content and designing the system interface.
   - Present information using multiple, complementary symbols, formats, and perspectives.
   - Embed activities that increase the perceived demand characteristics of both the media and learning activities.
   - Structure presentations and interactions to complement cognitive processes and reduce the complexity of processing tasks.
• Anchor knowledge in realistic contexts and settings.
• Provide methods that help learners acquire knowledge from multiple perspectives and cross-reference knowledge in multiple ways.
• Differentiate orienting activities for forthcoming information based upon desired learning; provide organising activities for information previously reviewed.
• Provide opportunities to respond and receive response-differentiated feedback where critical information is involved, but avoid excessive focus on responses whenever incidental learning is expected.
• Differentiate key terms, concepts, and principles through cosmetic amplification, repetition, and recasting.
• Provide clearly defined procedures for navigating within the system and accessing on-line support.
• Provide concept maps to indicate the interrelationships among concepts, and hypermaps to indicate the location of the learner relative to other lesson segments.
• Provide tactical, instructional, and procedural assistance.
• Interactive multimedia must adapt dynamically to both learner and content characteristics.
• Provide prompts and self-check activities to aid the learner in monitoring comprehension and adaption of individual learning strategies.
• Employ screen design and procedural conventions that require minimal cognitive resources, are familiar or can be readily understood, and are consistent with learning requirements.

The above guidelines can be further developed to form a heuristic evaluation instrument that assists in determining to what extent multimedia applications meet the twenty educational principles.

3. Barker and King (1993) developed a methodology for evaluating interactive multimedia courseware. The evaluation methodology involves twelve facets, which are:

• engagement
• interactivity
• tailorability
• appropriateness of multimedia mix
• mode and style of interaction
• quality of interaction
• quality of end-user interfaces
• learning styles
• monitoring and assessment techniques
• built-in intelligence
• adequacy of ancillary learning support tools
• suitability for single user/group/distributed use.

The above twelve facets were further developed into a series of evaluation questions for expert and novice users respectively. The two versions of questionnaires are suitable for self-evaluation as well as TELT pilots, and are ready-for-use.

4. Albion (1999) adapted previous studies’ models and developed a series of heuristics to evaluate the user interface, educational design, and content of an educational multimedia project:
**Interface design**
- Ensures visibility of system status: the software informs the user about what is going on through appropriate and timely feedback.
- Maximises match between the system and the real world: the software speaks the users' language rather than jargon. Information appears in a natural and logical order.
- Maximises user control and freedom: users are able to exit locations and undo mistakes.
- Maximises consistency and matches standards: users do not have to wonder whether different words, situations or actions mean the same thing. Common operating system standards are observed.
- Prevents errors: the design provides guidance which reduces the risk of user errors.
- Supports recognition rather than recall: objects, actions and options are visible. The user does not have to rely on memory. Information is visible or easily accessed whenever appropriate.
- Supports flexibility and efficiency of use: the software allows experienced users to use shortcuts and adjust settings to suit.
- Uses aesthetic and minimalist design: the software provides an appealing overall design and does not display irrelevant or infrequently used information.
- Helps users recognise, diagnose and recover from errors: error messages are expressed in plain language, clearly indicate the problem and recommend a solution.
- Provides help and documentation: the software provides appropriate online help and documentation that is easily accessed and related to the users' needs.

**Educational design**
- Clear goals and objectives: the software makes it clear to the learner what is to be accomplished and what will be gained from its use.
- Context meaningful to domain and learner: the activities in the software are situated in practice and will interest and engage a learner.
- Content clearly and multiply represented and multiply navigable: the message in the software is unambiguous. The software supports learner preferences for different access pathways. The learner is able to find relevant information while engaged in an activity.
- Activities scaffolded: the software provides support for learner activities to allow working within existing competence while encountering meaningful chunks of knowledge.
- Elicit learner understandings: the software requires learners to articulate their conceptual understandings as the basis for feedback.
- Formative evaluation: the software provides learners with constructive feedback on their endeavours.
- Performance should be 'criteria referenced': the software will produce clear and measurable outcomes that support competency-based evaluation.
- Support for transference and acquisition of 'self-learning' skills: the software supports transference of skills beyond the learning environment and facilitates self-improvement.
- Support for collaborative learning: the software provides opportunities and support for learning through interaction with others through discussion or other collaborative activities.

**Content design**
- Establishment of context: in a simulated reality, photographs, documents and other materials related to the simulated schools create a sense of immersion.
- Relevance to professional practice: the problem scenarios and included tasks are realistic and relevant to the professional practice of teachers.
• Representation of professional responses to issues: the sample solutions represent a realistic range of teacher responses to the issues and challenge users to consider alternative approaches.
• Relevance of reference materials: the reference materials included in the package are relevant to the problem scenarios and are at a level appropriate to the users.
• Presentation of video resources: the video clips of teacher interviews and class activities are relevant and readily accessible to the user.
• Assistance is supportive rather than prescriptive: the contextual help supports the user in locating relevant resources and dealing with the scenarios without restricting the scope of individual responses.
• Materials are engaging: the presentation style and content of the software encourages the user to continue working through the scenarios.
• Presentation of resources: the software presents useful resources for teacher professional development in an interesting and accessible manner.
• Overall effectiveness of materials: the materials are likely to be effective in increasing teachers’ confidence and capacity for integrating information technology into teaching and learning.

Albion’s (1999) heuristic list covers three important areas of educational multimedia design and development, and has theoretical foundations. These heuristics are ready-for-use for the TELT evaluation framework.

5. Pham (1998) maintained that multimedia systems should be evaluated from three perspectives: the application itself, how the application is used, and the impacts of the application. Followed an analysis of the characteristics of high quality multimedia systems, Pham (1998) developed a series of questions for evaluation:

• Knowledge content and tasks
  • Are the knowledge content and its organisation appropriate for achieving the specified objectives?
  • Are they pitched at the right complexity level for the users that the system attempts to reach?
  • Do the tasks that are designed to convey this knowledge stimulate and enhance users’ capacity for learning?
• Technical aspects
  • How responsive is the system for interactive use?
  • Is it fast enough to provide real-time responses, and if not, is there any message to inform users of what is going on?
  • What technical features are offered by the system?
  • How reliable are they under extensive use by different types of users?
  • Are these features better than those provided by similar systems?
  • Do they provide new functionalities or innovative way to perform a specific task?
  • Were these capabilities be implemented in such a way that can be extended easily?
• Human computer interface
  • How well can users carry out the targeted application using the interface?
  • What is the degree of users’ satisfaction?
  • How easy is it for users to learn ways to operate the system and remember them?
  • How effectively and efficiently can users perform each of these specific tasks?
• Navigation and links to information structures
How readily can users construct knowledge or perform tasks by following the links provided by the system?

Does the information content provided in each node and its associated nodes facilitate relational understanding of concepts?

How do such links and navigation methods provide more effective ways to disseminate knowledge than traditional media?

Do they stimulate creative ideas and commitment?

Impacts of application

How well can the application be deployed within the curriculum?

Has the application achieved the intended objectives?

Has the application performed well in comparison with similar products?

The questions that Pham (1998) raised may be treated as a general reference only because Pham’s (1998) thoughts have not been formed into a clear framework, nor did Pham (1998) actually use the proposed questions to evaluate multimedia applications. The questions were led by general discussions on different aspects of the quality of multimedia systems.

6. Kennedy et al., (1998) reviewed relevant literature and identified four main areas that may be used as a basis for devising criteria for evaluating educational multimedia applications:

- Theories and models of learning and instruction
- Instructional and conceptual design
- Navigation and orientation
- Interactivity
- Sequencing
- Consistency between learning objectives and content of instruction
- Interface and Graphic Design
- User attitudes and affect.

Kennedy et al., (1998) did not supply exact evaluation criteria. Instead, they justified the four areas, which were supported by research literature. The terms, scopes, and the essential meanings of the four areas articulated by Kennedy et al., (1998) are useful for developing criteria for the TELT framework.

7. Heller et al., (2001) developed a multimedia taxonomy for evaluation that highlight the following aspects:

- Evaluation dimensions
  - aesthetics
  - audience
  - discipline
  - interactivity
  - quality
  - usefulness
- Media types
  - text
  - sound
  - graphics
• motion
• multimedia
• Expression types
  • general
  • elaboration
  • representation
  • abstraction.

Within each evaluation dimension, there are questions for each media type and its corresponding expression type. Although not all areas within the framework included evaluation questions (some areas are left blank), the framework does cover most areas relevant to multimedia applications. This framework is useful for developing criteria for the evaluation of multimedia applications delivered through the TELT platform.

1.1.7 Web based lecture technologies

Web based lecture technologies (WBLT) allow automatic recording of live lectures and processing into a variety of streaming media formats. Students can later access lecture recordings on the web and review the delivered material at their own convenience.

Gosper, McNeill, Phillips, Preston, and Woo (2008a) included the following reasons to utilise web based lecture technologies:

• to support students who are unable to attend class;
• to provide a study tool for review and revision;
• to cater for individual learning strategies and styles;
• to supplement face to face lectures, but at a time and place of the student’s choosing;
• to accommodate student expectations regarding the digital delivery of course material; and
• to facilitate distance education as an alternate delivery mode

Evaluation of WBLT is largely based on students / staff surveys and log data.

Table 6: Overview of the Methodologies Reviewed in Section 1.1.7

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. von Konsky, Ivins &amp; Gribble (2009)</td>
<td>Various</td>
<td>Survey, student data (logs and such), case studies</td>
<td>This method should be modified to allow students for downloading the lectures (and possibly counting the number of downloads). Case studies could be used at the later stages of the evaluation to add rich details to the identified patterns.</td>
</tr>
</tbody>
</table>
2. **Gosper et al., (2007b)** | Various | Survey, selective interviews and case studies | The questionnaire and semi-structured interview questions are suitable for students and staff to complete. However, the questionnaire could be further developed to include social presence and participant-interface questions and also to include a graphic results panel.

3. **McElroy & Blount (2006)** | Various | Survey for students, reflective journals for staff | The questionnaire has similar to #2 items, and it is not very detailed. Reflective journals, however, could be used with a staff in order to gather different type of evaluative information.

4. **Albon (2004)** | Various | Questionnaire, open-ended interviews | The questionnaire is an early attempt to measure WBLT role in high education. As an early attempt it doesn’t present a high level of details (unlike studies #2 and #3)

1. Von Konsky, Ivins and Gribble (2009) reviewed both Australian and international experiences in evaluating students’ usage of WBLT that employed a mixed approach based on surveys, students’ logs analysis, and a lecture attendance analysis. However, their approach was somewhat limiting for students (as participating students pointed out) in that they were not permitted to download the audio track to listen to at a more convenient time. They were required to only attend the streaming format sessions (audio track combined with powerpoint slides), which may have impacted on usage patterns and thereby the research conclusions. The case studies highlighted the importance of students’ engagement and the variety of individual learning strategies used.

2. Gosper, McNeil, Woo, Phillips, Preston and Green (2007b) reviewed relevant literature and the practitioners’ experiences in evaluating WBLT, and identified several criteria that demonstrate the differences in students’ and staff perceptions of WBLT. These criteria may be used as a basis for evaluating WBLT applications:

   - perceptions of the effectiveness of WBLT
   - usage patterns
   - reasons for using WBLT
   - lecture attendance and WBLT usage
   - differential approaches to teaching / learning

Differential factors identified by Gosper et al., (2007b) could also serve as a basis for deriving a graphic evaluative model reflecting the alignment of the students’ and staff perceptions of WBLT, and showing how teaching and learning principles could help to improve this alignment.

3. McElroy and Blount (2006) developed a survey framework for a field of study in Finances. Their WBLT evaluation framework included the following themes:

   - students learning experiences
   - factors contributing to WBLT usage
   - reasons for using WBLT
   - comparison of f-2-f environment and WBLT
Although McElroy and Blount’s survey served its evaluative purpose, the survey items could be further developed to include more details. However, the reflective journals used in this evaluation could serve the purposes of a holistic evaluation and thus be useful for reflection on the opinions of different stakeholder groups.

4. Albon (2004) examined the pedagogical value of ilectures by using a holistic approach to educational delivery and subscribing to constructivist learning theory. Her survey covered several emerging themes in ilecture use. In particular, the survey questions were presented under the following sections:

- who uses ilectures
- when are ilectures used
- how are ilectures used
- why are ilectures used or not used.

Survey questions are similar to those used in the Gosper et al., (2007b) study, but do not cover approaches to teaching and learning. The questionnaire is an early (basic) version of an evaluation technique commonly used for WBLT.

At this stage of the evaluation framework development process, the decision to adopt WBLT is not in question, but rather the capacity to repurpose WBLT tools, adapt them to other campus technologies, and to use them “effectively as study tools to enhance student learning” (Gosper et al., 2007b, p. 5). The current framework will incorporate the findings of previous studies such as McNeill, Woo, Gosper, Phillips, Preston and Green (2007b), in particular their advice regarding: a) the structure and the content of the lecture; b) the lecturing process and c) managing the technical aspects of WBLT, which in turn provide new methods in assessing the educational value of WBLT applications.

Finally, as explained in detail in the document titled “MindMap3.doc”, lecture capture applications such as Lectopia or Echo360 should not be viewed in the same light as Educational Multimedia Software. The main distinction lies in the recording of actual class sessions that are used to augment classroom instruction as opposed to the design and development of separate educational content that are not derived from students’ prior direct experiences.

1.1.8 Educational Value

Many of the methodologies outlined in the preceding sections are concerned with the usability of particular types of software applications. Although the results of usability evaluation may indicate the degree of the selected application’s educational value, there are methodologies that target the educational value directly as they reflect two important components of educational value: (1) the integration of knowledge in the learner, and (2), the integrated development of learners (Crittenden, 1968).

In the context of technology-enabled learning and teaching, if an eLearning application assists educators to achieve educational value effectively and efficiently, then it can be
recognised as providing high educational value. Although the following literature does not specify specific types of eLearning software applications to be evaluated, the methodologies described manifestly reflect an educational value perspective.

Jackson (1998) argues that the learning outcomes obtained through the implementation of an eLearning package should be measured against the following criteria:

- **Ease of use**: can the target users navigate and interact with the application in the expected way?
- **Efficiency**: does the application provide efficient coverage of the appropriate curricular unit? Can more learners access the content than by other means?
- **Preferences**: do learners enjoy it? will they choose to use it if it is offered? to what extent does it cater for a variety of users?
- **Attractiveness**: is it sufficiently attractive to persuade purchasers/tutors to choose it over alternatives?
- **Cost effectiveness**: is it cost effective in comparison with alternative means of achieving the same intended outcomes?

Jackson (1998) further proposed that the SOLO taxonomy (Biggs, 1996) can be used to measure eLearning outcomes and to determine whether the software application provides opportunities for learning at an appropriate level. The SOLO taxonomy describes five levels of understanding that can be encountered in learners’ responses to academic tasks:

- **Prestructural**: the task is not approached appropriately (the student has not understood the point)
- **Unistructural**: one or a few aspects of the task are picked up and used (understanding is nominal)
- **Multi-structural**: several aspects of the task are learned, but are treated separately (understanding is knowing about)
- **Relational**: the components are integrated into a coherent whole, with each part contributing to the overall meaning (understanding is appreciating relationships)
- **Extended abstract**: the integrated whole at the relational level is reconceptualised at a higher level of abstraction, which enables generalisation to a new topic or area, or is turned reflexively on oneself (understanding is for knowledge transfer and as involves metacognition).

Jackson’s (1998) work is highly theoretical and has not been developed into an instrument, but it nevertheless useful for devising evaluation approaches and instruments that target learning activities that are facilitated by eLearning software applications.

The eLearning Evaluation Heuristics advanced by Reeves and Hedberg (2008) inform twenty aspects of an eLearning program. The first ten are based directly on Nielsen’s (1994a) ten heuristics for usable design:

- visibility of system status
- compatibility between the system and the real world
- user control and freedom
- consistency and standards
- error prevention
- recognition rather than recall
- flexibility and efficiency of use
• aesthetic and minimalist design
• help users recognise, diagnose, and recover from errors
• help and documentation
• interactivity
• message design
• learning design
• assessment
• media integration
• resources
• performance support tools
• learning management
• feedback
• content.

Each of aspects listed above contains three sample questions. All questions reflect instructional contexts are suitable for evaluating the educational value of the selected applications. Some of the sample questions include:

• Does the learner know where they are in the eLearning programme at all times, how they arrived there, and how to return to the point from which they started?
• Does the eLearning programme take advantage of state-of-the-art of eLearning design to engage the learner to complete content-specific tasks and solve problems?
• Does the eLearning programme provide opportunities for self-assessment activities that advance learner achievement?

Reeves and Hedberg (2003) also developed a range of eLearning evaluation templates and samples that can be used to manage the evaluation lifecycle, which are available for download at: http://is.gd/4F1CN

Shank (2004) measured the value that eLearning brings to organisations using a series of evaluation criteria design for that purpose. The following criteria are relevant to the higher education context:

• learning gains
• learner performance measures
• course enrolments
• course completion rates
• course revenue or profits
• student satisfaction with specific course
• student satisfaction with curriculum offerings
• increased student demand for eLearning
• reduction in teaching time
• reduction in teaching costs.

Shank (2004) also specified the data sources and evaluation methods for each of the above criteria above (for example, testing, LMS data tracking, surveys). In essence, they are broad criteria for evaluating educational value. To utilise these criteria, the sub-criteria and their evaluation context for use must be specified.
1.2 Learning Effectiveness of TELT Technologies

The majority of the methodologies used for evaluation of eLearning effectiveness either measure directly by testing and/or administering a questionnaire to students, or expert review and examination of other factors of the learning environment that reflect or predict the learning effectiveness. Some of the documents reviewed only provide conceptual models or theoretical arguments and are devoid of instruments that may assist to develop effective methodologies. Most documents reviewed in this section do not specify the types of eLearning software applications that were evaluated, but it is obvious that they were designed for learning management systems.

The goal in this section is to identify methodologies for the evaluation of eLearning effectiveness. Although there are a number of comparative studies that support the view that eLearning is at least as effective as traditional learning methods, but not more effective (Macgregor and Turner, 2009), the methodologies and related studies described in these studies have been excluded from this review.

1.2.1 Direct Evaluation of Learning Effectiveness

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework Stage 3, sub-layer 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shank (2004)</td>
<td>Various</td>
<td>N/A</td>
<td>The criteria proposed by Shank provided a starting point to develop evaluation instruments.</td>
</tr>
<tr>
<td>Figueira (2003)</td>
<td>Various Evaluation</td>
<td>Evaluation questions</td>
<td>Figueira’s work may serve as a general reference for the TELT Evaluation Framework and could be adapted to other evaluation models and/or instruments.</td>
</tr>
<tr>
<td>Pogarcic &amp; Vujic (2007)</td>
<td>Student Assessment</td>
<td>N/A</td>
<td>Pogarcic &amp; Vujic’s (2007) argument about eTaxonomy may be useful to specify evaluation approaches that correspond different levels of learning goals with the use of TELT technologies.</td>
</tr>
<tr>
<td>Dureva &amp; Tuparov (2006)</td>
<td>Student Assessment</td>
<td>Assessment Model</td>
<td>To apply Dureva &amp; Tuparov’s (2006) model and to take Pogarcic &amp; Vujic’s (2007) views into account, it is possible to create a reference system to link these elements together: levels of learning goals, assessment approaches, e-assessment tools or tools that may be used for assessment, eLearning applications, and eLearning product types.</td>
</tr>
</tbody>
</table>
Survey  
Questionnaire  

Liaw’s (2008) questionnaire covers various important aspects of eLearning evaluation and the analysis approach (e.g., figuring out the factors that can predict eLearning effectiveness) is worthwhile for us to learn from.

Expert Review  
Question Pool  

Two sections of the question pool are suitable for evaluating the learning effectiveness of TELT technologies.

1. There are several methodologies that are used for evaluating selected eLearning software applications that are also suitable for evaluation of the learning effectiveness of TELT Technologies as a whole. For example, Shank (2004) lists a series of criteria for evaluating educational value that also indicate the learning effectiveness attributable a range of educational technologies. Refer to Section 1.1.7 for a more detailed description of Shank’s (2004) criteria. The criteria that specifically indicate learning effectiveness are:

- Learning gains
- Examination and test scores
- Course completion rates
- Student satisfaction

2. Figueira (2003) listed five approaches that cater to the different reasons for evaluating eLearning effectiveness, which were adapted from Pietro (1983):

<table>
<thead>
<tr>
<th>Model</th>
<th>Main Aim</th>
<th>Typical Questions</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programme Goals</td>
<td>Estimating how goals have been reached</td>
<td>Have goals been reached? Have they been effectively and efficiently reached? Were goals appropriate?</td>
<td>Quantitative. Experimental design. Before and after testing</td>
</tr>
<tr>
<td>Goal-free</td>
<td>Estimating efforts developed by the programme</td>
<td>What are the intentional and non-intentional results of the programme? How much value has the programme resulted in?</td>
<td>Independent estimation from needs and judging criteria. Quantitative and qualitative techniques.</td>
</tr>
<tr>
<td>Expert</td>
<td>Experts as judges and evaluation tools</td>
<td>How an external professional classifies the programme?</td>
<td>Critical revision based on experience and collection of specific information and subjective opinions.</td>
</tr>
<tr>
<td>Naturalistic</td>
<td>Understanding the process behind the programme</td>
<td>What is happening to programme? What are the different perspectives? How is the programme responding to the different interests?</td>
<td>Inductive discovery based on qualitative methods. Open interviews, participant observation and case studies.</td>
</tr>
</tbody>
</table>
Figueira (2003) further proposed a global approach that provides a mix of decision-making, goal-free, and expert evaluation models. This approach points to four levels of evidence for determining learning effectiveness:

- resources
- participation
- reaction
- results.

To make the global approach practical, Figueira (2003) developed a list of evaluation criteria that are grouped into the following categories:

- organisational and management dimensions
- pedagogical dimension
- technological dimension
- ethical considerations
- learning assessment and certification
- evaluation strategy.

Figueira (2003) claimed that the three main outcomes (that is, the table, the level of evidence, and the criteria) were linked together and that the third is dependent on the first two. However, the links are not obvious, nor did Figueira articulate clearly how they were developed and how they could be used for eLearning effectiveness evaluation. However, Figueira’s work could serve as a general reference for the TELT Evaluation Framework and potentially adapted to other evaluation models and/or instruments.

3. Spitzer (2005) developed a Learning Effective Measurement (LEM) methodology to address the challenge of providing a credible result-oriented learning measurement that would not only help to evaluate learning interventions, but also increase their effectiveness. The LEM methodology includes five phases:

- predictive measurement
- baseline measurement
- formative measurement
- in-process measurement
- retrospective measurement.

Spitzer’s (2005) methodology progresses through the whole process of a typical eLearning development and implementation cycle. However, the LEM methodology is more accurately a conceptual model and does not provide any instruments apart from work sheets for the first phase. It informs what to do at a particular time of the cycle, but it does not detail the procedures for each phase.

4. Pogarcic and Vujic (2007) acknowledge that Bloom’s taxonomy is applicable to an assessment of the level of goals achieved through instruction in traditional settings, but claim that the application of ICT has changed various elements of instruction. Although elements contained in the traditional assessment taxonomy may have relevance for the eLearning context, Pogarcic and Vujic did not propose to adapt a new taxonomy from
Bloom’s taxonomy. Nonetheless, they maintain that in the eLearning context, three of the elements should be taken into account:

- Collaboration
- Assignments, problem solving and projects
- Discussion

Pogarcic and Vujic (2007) did not specify what exactly should be changed in the eLearning context, but they stress that changes in key terms for eLearning evaluation are dependent on levels, instruction contents, age and capabilities of students, their existing knowledge, capabilities and skills of teachers, and technological possibilities (the application to a LMS, the method for storing and organising learning resources within the LMS, communication, and moderation). Pogarcic and Vujic’s (2007) general discussion does not provide direction for change or a theoretical foundation. However, in the TELT Evaluation Framework context, Pogarcic and Vujic’s views on developing a new taxonomy or modifying Bloom’s taxonomy may assist in the evaluation of learning effectiveness of TELT technologies. For example, different evaluation approaches may be specified for each of the corresponding levels of learning goals that are intended through the use of TELT technologies.

5. Dureva and Tuparov (2006) proposed a model for linking the common assessment approaches used in eLearning contexts to the tools that may generally be available in eLearning environments.

**Figure 2: Common assessment approaches and evaluative tools**

In order to apply Dureva and Tuparov’s (2006) model to the TELT Evaluation Framework and to take advantage of Pogarcic and Vujic’s (2007) perspectives, a reference system can be constructed to link the following elements to form a useful assessment model for eLearning activities:

- levels of learning goals
- assessment approaches
e-assessment tools or tools that may be used for assessment
eLearning applications
eLearning application types.

6. Liaw (2008) conducted a research study to identify the relationships amongst eLearning effectiveness, students’ perceived satisfaction, and behavioural intention. Four hundred and twenty four student users of Blackboard from a university in Taiwan completed a questionnaire that comprised the following Likert scale questions:

- **Perceived self-efficacy**
  - I feel confident using the e-leaming system (Blackboard)
  - I feel confident operating eLearning functions
  - I feel confident using online learning contents

- **Perceived satisfaction**
  - I am satisfied with using eLearning as a learning assisted tool
  - I am satisfied with using eLearning functions
  - I am satisfied with learning contents
  - I am satisfied with multimedia instruction

- **Perceived usefulness**
  - I believe eLearning contents are informative
  - I believe eLearning is a useful learning tool
  - I believe eLearning contents are useful Behavioural intention
  - I intend to use eLearning to assist my learning
  - I intend to use eLearning content to assist my learning
  - I intend to use eLearning as an autonomous learning tool eLearning system quality
  - I am satisfied with eLearning functions
  - I am satisfied the Internet speed
  - I am satisfied with eLearning content
  - I am satisfied with eLearning interaction Interactive learning activities
  - I would like to share my eLearning experience
  - I believe eLearning can assist teacher-learner interaction
  - I believe eLearning can assist learner-learner interaction

- **ELearning effectiveness**
  - I believe eLearning can assist learning efficiency
  - I believe eLearning can assist learning performance
  - I believe eLearning can assist learning motivation

- **Multimedia instruction**
  - I like to use voice media instruction
  - I like to use video media instruction
  - I like to use multimedia instruction

After a series of regression and correlation analysis, Liaw (2008) applied the following model to demonstrate their results.
The above model indicates that eLearning effectiveness can be predicted by three independent factors: interactive learning activities; multimedia instruction; and eLearning system quality. Multimedia was the biggest predictor and contributed 59.3% to eLearning effectiveness. Behavioural intention for using eLearning thus correlates with eLearning effectiveness.

Liaw’s (2008) questionnaire covers various important aspects of eLearning evaluation. The applied analysis approach is highly noteworthy for further application to the TELT Platform. However, the number of questions for investigating students’ perceived eLearning effectiveness appeared limited and this restriction was not justified by the author or other sources. Therefore, while the format of Liaw’s (2008) evaluation is useful for the development of the TELT Evaluation Framework, the variables for eLearning effectiveness may need to be expanded.

7. Lam and McNaught (2004) developed a strategy to evaluate thirty educational websites designed by lecturers at three universities in Hong Kong. The strategy came with an evaluation question pool that contained two hundred and ninety items organised under twenty nine headings. Two sections of the question pool are Teaching and Learning Processes and Learning Outcomes, both of which are suitable for evaluating the learning effectiveness of TELT technologies. The questions for Teaching and Learning Processes can be accessed from:

http://e3learning.edc.polyu.edu.hk/Qdb-T.htm

and the questions for Learning Outcomes are available at:

1.2.2 Indirect Evaluation of Learning Effectiveness

Table 9: Overview of the Methodologies Reviewed in Section 1.2.2

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework Stage 3, sub-layer 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonwalkar (2002)</td>
<td>Score Rating</td>
<td>Pedagogy Effectiveness Index (PEI)</td>
<td>Sonwalkar’s (2002) methodology utilises the relevant objective facts to indicate eLearning effectiveness. It is necessary to set up the thresholds for different PEI results and overall rating results, and allocate various implications to the thresholds so that the formula-generated figures can be more meaningful.</td>
</tr>
<tr>
<td>Schank (2002)</td>
<td>N/A</td>
<td>FREEDOM Model (Criteria List)</td>
<td>The FREEDOM model provides seven important criteria for effective eLearning and can be used for the development of a more formal evaluation instrument.</td>
</tr>
<tr>
<td>Macgregor &amp; Turner (2009)</td>
<td>N/A</td>
<td>Conceptual model – list of factors that have great impacts on eLearning effectiveness</td>
<td>The model has strong support from previous studies. It can be further developed into an evaluation instrument for the TELT Evaluation Framework.</td>
</tr>
</tbody>
</table>

1. Sonwalkar (2002) proposed an instrument for eLearning effectiveness evaluation, based on a five-factor summative rating system plus a pedagogy effectiveness index (PEI). The intent of the methodology was to create objective criteria for evaluating the quality of online courses based on existing elements that represent pedagogical content. According to Sonwalkar (2002), eLearning effectiveness can be defined as a summation of up to 5 learning styles, up to 6 media elements, and up to 5 interactivities. The pedagogy effectiveness index can be calculated using the following formula:

\[ PEI = \sum S_i \cdot p_i + \sum M_j \cdot p_j + \sum I_k \cdot p_k \]

Where \( S = \) Style, \( M = \) Media, and \( I = \) Interaction; the subscripts define the elements’ ranges: \( i = 1 \) to \( 5 \), \( j = 1 \) to \( 6 \), and \( k = 1 \) to \( 5 \); and \( \sum \) represents summation.

The probability distribution used in the formula is as followed.

Table 10: Weights and distribution of different pedagogical elements for PEI

<table>
<thead>
<tr>
<th>Style</th>
<th>( p_i )</th>
<th>Media</th>
<th>( p_j )</th>
<th>Interaction</th>
<th>( p_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apprenticeship</td>
<td>0.068</td>
<td>Text</td>
<td>0.055</td>
<td>Feedback</td>
<td>0.066</td>
</tr>
<tr>
<td>Incidental</td>
<td>0.068</td>
<td>Graphics</td>
<td>0.055</td>
<td>Revision</td>
<td>0.066</td>
</tr>
<tr>
<td>Inductive</td>
<td>0.068</td>
<td>Audio</td>
<td>0.055</td>
<td>E-mail</td>
<td>0.066</td>
</tr>
<tr>
<td>Deductive</td>
<td>0.068</td>
<td>Video</td>
<td>0.055</td>
<td>Discussion</td>
<td>0.066</td>
</tr>
<tr>
<td>Discovery</td>
<td>0.068</td>
<td>Animation</td>
<td>0.055</td>
<td>Bulletin</td>
<td>0.066</td>
</tr>
</tbody>
</table>
For example, the PEI for a course with three learning styles, four media elements, and two interactive elements will be: PEI = 3 * 0.068 + 4 * 0.055 + 2 * 0.066 = 0.556. The probability of the pedagogical effectiveness increases as cognitive opportunity increases with the inclusion of learning styles, media elements, and interaction. Besides the PEI indicator, Sonwalkar (2002) also incorporate additional factors to the methodology and developed a list of objective criteria for a summative evaluation in five major areas:

- content factors
- learning factors
- delivery support factors
- usability factors
- technological factors.

The summative evaluation results (the sum of the ratings of all the factors in each of the five categories) and the PEI can be combined to give a final result that provides a view of the overall effectiveness of the online course:

\[
\text{Overall Rating} = \text{PEI} \times \text{Summative Rating Score}
\]

Sonwalkar’s (2002) methodology utilises the relevant objective facts to indicate the eLearning effectiveness, but no explanation as to how the methodology was developed and how the instruments were justified, if ever. Users of the methodology may find it difficult to interpret the numeric results as there is no interpretation mechanism available. Therefore, to make full use of Sonwalkar’s (2002) methodology, it will be necessary to establish thresholds for different PEI results and overall rating results, and allocate different implications to the thresholds so that the formula-generated figures can provide more meaningful results.

2. Schank (2002) developed a model called FREEDOM that comprises seven means by which to evaluate the effectiveness of an eLearning program, namely:

- Failure: enables students to learn from their mistakes
- Reasoning: encourages practice in deliberating decisions
- Emotionality and Exploration: provide a more engaging environment for learners by allowing them to form an emotional link to material while also having the option to inquire or further discuss a topic
- Doing: involves the provision of training for students to apply their knowledge in real-life situations
- Observation: all course materials should be presented in a user-friendly format and supported with the opportunity for observation which includes the provision of diagrams charts and other visual aids
- Motivation: provides the student with a feeling of being able to personally relate to the material and its value.
The FREEDOM model provides seven important criteria for determining effective eLearning which can be used for the development of a more formal evaluation instrument.

3. Macgregor and Turner (2009) conducted a comprehensive literature review about eLearning effectiveness and concluded that the following factors exert significant impact on eLearning effectiveness:

- External factors
  - ICT literacy
  - information literacy
  - learning styles
  - assessment strategy
  - learning experience
- Internal factors
  - system design and usability
  - interactivity
  - information architecture
  - learning object quality
  - teacher/learning technologist
- Link between external and internal factors
  - learner control.

This is a conceptual model that involves themes that emerged form related research literature. In other words, the model has strong support from previous studies. It can be further developed into an evaluation instrument for the TELT Evaluation Framework.

1.3 Learning Environments that Use TELT Technologies

<table>
<thead>
<tr>
<th>Document</th>
<th>Methodology</th>
<th>Available Instrument</th>
<th>Relevance to the TELT Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Attwell (2006)</td>
<td>Heuristic Evaluation</td>
<td>Heuristics</td>
<td>Attwell’s (2006) evaluation questions were originally designed for evaluating policies. The criteria are useful for evaluating the design rationale of eLearning courses built into an eLearning environment. The main drawback of the instrument is that it may be difficult for the evaluators to determine whether the policy is meeting the requirements in the questions due to lack of indicators in the instrument. If this</td>
</tr>
</tbody>
</table>
instrument is adapted, it will also assist evaluators if the evaluation questions included examples.

4. Agostinho et al., (2002)  Expert Review  Evaluation Form  Agostinho et al., (2002) also specified the suggested usage of evaluation forms for different purposes—formative evaluation or summative evaluation. This heuristic evaluation instrument is ready-for-use and can be adapted to the TELT Evaluation Framework.


7. Reeves (1994)  N/A  Theoretical model  To utilise Reeves’s (1994) model, it is possible to change the number of dimensions to accord with different evaluation purposes and use the refined model to gain a better understanding of eLearning environment designs in terms of their pedagogical foundations. However, in the context of the TELT Evaluation Framework, Reeves’s (1994) model does not help to judge which technology-enabled learning environments are more effective unless there are other evaluation criteria assigned each pedagogical dimension so that the ratings provide more meaningful evaluation implications. Therefore, Reeves’s (1994) model needs to be combined with other methodologies or used to develop a TELT methodology.

1. As noted in Section 1.2.1, Lam and McNaught (2004) developed a strategy to evaluate thirty educational websites that included an evaluation question pool, which contains two hundred and ninety items organised under twenty nine headings. While two sections of the question pool target evaluation of learning effectiveness, the remainder targets various aspects of a typical eLearning environment. Therefore, the questions appear suitable for evaluating learning environments that use TELT technologies. The evaluation questions developed by Lam and McNaught (2004) are grouped under the following headings:

- Pre-developmental (P)
- expectation
- habits
- motivation
- Environmental (E)
- appreciation
2. Grant MacEwan College in Canada released a set of evaluation guidelines to assist educators in evaluating online courses (Wright, 2004). Each evaluation item identifies specific course components, qualities, or procedures proven to be helpful to learners and/or instructors. There are a total of one hundred and thirty nine items grouped into the following categories:

- General information
- Accessibility
- Organisation
- Language
- Layout
- Goals and Objectives
- Course Content
- Instructional or Learning Strategies and Opportunities for Practice and Transfer
- Learning resources
- Evaluation
Overall

These criteria cover most aspects of an eLearning environment and are directly useful for development of the TELT Evaluation Framework.

3. Attwell (2006) developed a list of questions for evaluating policy for eLearning environments. The criteria are concerned with the extent to which the policy has considered individuals’ various characteristics. The following are the main questions to be resolved:

- Does the policy take account of the fact that the physical characteristics of individuals impacts on their behaviour as eLearners?
  - Does the policy recognise that:
    - age makes a difference?
    - gender makes a difference?
    - physical disability / ability makes a difference?
- Does the policy take account of the fact that the learning history of individuals impacts on their behaviour as e-learners?
  - Does the policy recognise that:
    - learners’ previous level of attainment makes a difference?
    - the quantity / duration of the learners’ previous learning makes a difference?
    - The learners’ response to and experience of previous learning makes a difference?
    - the frequency or recentness of their learning experiences makes a difference?
- Does the policy take account of the fact that the attitude and motivation of the learner impacts on their eLearning behaviour?
  - Does the policy recognise that:
    - learners’ reasons for undertaking eLearning makes a difference?
    - learners’ expectations of an eLearning experience make a difference?
    - learners’ perceptions of eLearning make a difference
    - learners’ commitment and application make a difference?
- Does the policy take account of the fact that Learners familiarity with the technology and the learning environment impacts on their eLearning behaviour?
  - Does the policy recognise that:
    - the learners’ existing competence in the use of technology makes a difference?
    - the learners familiarity with the technology makes a difference?
    - whether the learners have a positive or negative attitude towards the technology makes a difference?
    - whether the learners have previous experience of eLearning makes a difference?

Although Attwell’s (2006) evaluation questions were originally designed for evaluating policies, the criteria are also useful for evaluating the design rationale of eLearning courses built into eLearning environments. The main drawback of the instrument is that it may be difficult for evaluators to determine whether the policy meets the requirements raised in the questions due to a lack of indicators within the instrument. If the instrument is to be adapted to the TELT Evaluation Framework, it will be helpful for evaluators if the questions included examples.
4. Agostinho et al., (2002) suggest that the design of eLearning environments should include three critical elements: learning tasks, learning supports and learning resources (the figure below illustrates the relationship among the three elements)

**Figure 4: Design of eLearning environments model**

Based on this three-element model, Agostinho et al., (2002) developed an evaluation form for expert review. The main questions in the form are listed below (additional details for each question are provided in the actual instrument):

- How does the learning design support Learner Engagement?
- How well does the learning design support Learner Engagement?
- How does the learning design acknowledge the learning context?
- How well does the learning design acknowledge the learning context?
- How does the learning design seek to challenge learners?
- How well does the learning design challenge learners?
- How does the learning design provide practice?
- How well does the learning design provide practice?
- Infrastructure and Technology assessment: How do the technologies employed, their supportive systems and particular implementation facilitate the learning design?
- Description of the Learning Design
- Summary description of the learning design
- Suitability for Redevelopment.

Agostinho et al., (2002) also specified the suggested use of the form for different purposes (formative or summative evaluation). This heuristic evaluation instrument is ready-for-use and can be adapted to the TELT Evaluation Framework.
5. Boud and Prosser (2002) developed a theoretical framework for the analysis of learning designs using new technologies. It takes a learner-centred view derived from literature in higher, professional and adult education. The framework consists of four key areas:

- Engaging learners. This includes starting from where learners are, taking into account their prior knowledge and desires, and building on their expectations.
- Acknowledging the learning context. This includes the context of the learner, the course of which the activity is part and the sites of application of the knowledge being learned.
- Challenging learners. This includes attempting to encourage learners to be active in their participation, using the support and stimulation of other learners, taking a critical approach to the materials, and to go beyond what is immediately provided.
- Providing practice. This includes demonstration of what is being learned, gaining feedback, reflection on learning, and developing confidence through practice.

Boud and Prosser (2002) used the following figure to portray the relationship of the four areas.

**Figure 5: Relationship of different elements in creation of high quality learning activities**

6. Admiraal, Lam, and Jong (2006) suggest three models for the evaluation of technology-enabled learning environments. The three models assist evaluators to gain a better understanding on whether the use of technologies confirms the tenets or principles described in educational theories.

- Model of learning processes in virtual learning environments:
  - *From guided learning to experiential and action learning: Admiraal et al. (2005)* suggest that eLearning environments should possess the following factors to ensure the success of transforming traditional guided learning to experiential and action learning:
- Interests, knowledge and action-plans of participants are put central
- The experience itself is the goal. There are no explicit learning goals
- Each learner can have his/her own tempo
- Team learning from and with each other is important
- Reward and judgment systems are tuned to discoveries and innovations
- Control of learning by learners
- Opportunities to reflect on learning goals and learning strategies
- Self-responsibility for their own learning
- Opportunity for self-testing

- Model of teachers’ pedagogy in virtual learning environments: Admiraal et al. (2005) suggest that teachers’ pedagogy in virtual learning environments is reflected in the ways they moderate computer-mediated communication. Admiraal et al. (2005) developed the following table to aid in the development of instruments for evaluating the moderation of eLearning environments with an emphasis on three major functions of moderators—organisational, social and intellectual.

### Table 12: Model of teachers’ pedagogy in VLEs

<table>
<thead>
<tr>
<th>Moderator Role</th>
<th>Organisational</th>
<th>Social</th>
<th>Intellectual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setter</td>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discriminator</td>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Host</td>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pace setter</td>
<td>high</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Explainer</td>
<td></td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Entertainer</td>
<td></td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>low</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Tutor</td>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitator</td>
<td>high</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>Mediator</td>
<td>high</td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Mentor</td>
<td>low</td>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Provocateur</td>
<td></td>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Observer</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>Participant</td>
<td>low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Consumer model for teachers in virtual learning environments: Admiraal et al. (2005) did not describe this model in detail, but they suggest that the information from end users should be used in the analysis of the functionalities of learning environments.

Admiraal et al. (2005) claimed that they developed two tools based on these models, but neither are available in the document reviewed. Despite this, Admiraal et al.’s (2005) discussion about the models may be of use for the development of TELT instruments. Reeves (1994) developed a meta-model for the design and evaluation of eLearning environments, highlighting 14 pedagogical dimensions. The fourteen dimensions form the basis for comparison of different learning environment designs:

- Epistemology: Objectivism ↔ Constructivism
- Pedagogical philosophy: Instructivist ↔ Constructivist
- Underlying psychology: Behavioural ↔ Cognitive
- Goal orientation: Sharply-focused ↔ Unfocused

86
Quinton, Dreher, Fisher, and Houghton (2005) extended Reeve’s (1994) model specifically for the design of object-based learning environments. The inclusion of the new continua extend Reeves’ model of evaluative pedagogical comparison to incorporate a framework for developing advanced online educational learning environments. The additional dimensions are:

- Technological complexity: Simple ↔ Complex
- Navigational control: Fixed ↔ Dynamic
- Object re-usability: Low ↔ High
- Object granularity: Low ↔ High
- Object contextualisation: High ↔ Low
- Object adaptivity: Low ↔ High
- Object scalability: Low ↔ High
- Object skill/knowledge specialisation: Low ↔ High
- Responsibility for assessment: Teacher ↔ Learner

Granić, Mifsud, and Ćukušić (2009) used Reeves’s (1994) model to evaluate multiple learning environments. The procedure was that the learning scenarios were rated by teachers (self evaluation), as well as four expert pedagogical evaluators who observed firsthand the implementation of the scenarios in secondary schools. The evaluation was performed using a form with a sliding scale for every dimension of the scenarios being studied. The outcome of each dimension was estimated qualitatively (sliding scale) so as not to burden the user numerical scale. While analysing the results, the qualitative estimates were converted into numerical measurements as distances of the marked positions on the line (in centimetres), from 0 to 14. Below are a table and a figure representing Granić et al.’s (2009) evaluation results.
Table 13: Granić et al.’s (2009) evaluation results

<table>
<thead>
<tr>
<th>Scenario title</th>
<th>E-1</th>
<th>E-2</th>
<th>E-3</th>
<th>E-4</th>
<th>E-5</th>
<th>E-6</th>
<th>E-7</th>
<th>E-8</th>
<th>E-9</th>
<th>E-10</th>
<th>E-11</th>
<th>E-12</th>
<th>E-13</th>
<th>E-14</th>
<th>Scenario mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>The human being: nutrition and digestion</td>
<td>6.0</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>4.8</td>
<td>11</td>
<td>6.8</td>
<td>11</td>
<td>6.5</td>
<td>2.5</td>
<td>4.1</td>
<td>3.9</td>
<td>3.6</td>
<td>6.5</td>
<td>7.05</td>
</tr>
<tr>
<td>Ecosystem</td>
<td>9.7</td>
<td>10</td>
<td>9.7</td>
<td>10.8</td>
<td>11.6</td>
<td>11.2</td>
<td>9.8</td>
<td>10.2</td>
<td>10.2</td>
<td>9.2</td>
<td>8.7</td>
<td>12.4</td>
<td>10.5</td>
<td>9.3</td>
<td>10.24</td>
</tr>
<tr>
<td>Investigating global warming</td>
<td>8.6</td>
<td>7.3</td>
<td>8.7</td>
<td>8.2</td>
<td>8.2</td>
<td>7.6</td>
<td>9.6</td>
<td>9.2</td>
<td>7.8</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>9.5</td>
<td>6.7</td>
<td>8.34</td>
</tr>
<tr>
<td>Historical heritage of Tribolje</td>
<td>8.2</td>
<td>8.8</td>
<td>8.7</td>
<td>9.5</td>
<td>8.6</td>
<td>8</td>
<td>6.3</td>
<td>8</td>
<td>8.3</td>
<td>9.3</td>
<td>8</td>
<td>8.7</td>
<td>12</td>
<td>1</td>
<td>8.10</td>
</tr>
<tr>
<td>Traffic Survey</td>
<td>8.8</td>
<td>7.5</td>
<td>9.1</td>
<td>8.4</td>
<td>8</td>
<td>7.7</td>
<td>9.7</td>
<td>10.3</td>
<td>7.7</td>
<td>7</td>
<td>6.8</td>
<td>10.2</td>
<td>9.6</td>
<td>7</td>
<td>8.41</td>
</tr>
<tr>
<td>Environmental Studies</td>
<td>9.7</td>
<td>10.2</td>
<td>8.7</td>
<td>10.3</td>
<td>11.6</td>
<td>10.5</td>
<td>9</td>
<td>11.6</td>
<td>9</td>
<td>7</td>
<td>9.6</td>
<td>11.2</td>
<td>9</td>
<td>7</td>
<td>9.68</td>
</tr>
<tr>
<td>West Galilean freshmen</td>
<td>10.5</td>
<td>9.7</td>
<td>9</td>
<td>5.6</td>
<td>10.5</td>
<td>6.6</td>
<td>7.9</td>
<td>5.7</td>
<td>12</td>
<td>7.6</td>
<td>6.8</td>
<td>9.6</td>
<td>8.2</td>
<td>7</td>
<td>8.41</td>
</tr>
<tr>
<td>Ancient Agora of Athens</td>
<td>13</td>
<td>13</td>
<td>12.9</td>
<td>12.6</td>
<td>12.5</td>
<td>12.5</td>
<td>10.6</td>
<td>13.5</td>
<td>9.7</td>
<td>10.8</td>
<td>8.5</td>
<td>8.2</td>
<td>11.2</td>
<td>1</td>
<td>10.64</td>
</tr>
<tr>
<td>Creating databases</td>
<td>3</td>
<td>3</td>
<td>3.8</td>
<td>4</td>
<td>11.2</td>
<td>11.2</td>
<td>7.2</td>
<td>3</td>
<td>2.4</td>
<td>4.5</td>
<td>11.3</td>
<td>3.6</td>
<td>7.2</td>
<td>5.8</td>
<td>8.88</td>
</tr>
<tr>
<td>New ICTs – visiting BALT Expo 2007</td>
<td>7.8</td>
<td>7.6</td>
<td>7.7</td>
<td>8.5</td>
<td>12</td>
<td>7.8</td>
<td>7.8</td>
<td>10.8</td>
<td>9</td>
<td>7</td>
<td>13</td>
<td>10.2</td>
<td>7.2</td>
<td>8.89</td>
<td></td>
</tr>
<tr>
<td>ICT in education – guide of Tribolje city</td>
<td>8.4</td>
<td>8.2</td>
<td>8.4</td>
<td>8.2</td>
<td>7.8</td>
<td>8.3</td>
<td>6.8</td>
<td>8</td>
<td>8.8</td>
<td>7</td>
<td>8</td>
<td>8.4</td>
<td>12.4</td>
<td>1</td>
<td>7.64</td>
</tr>
<tr>
<td>Youth crime</td>
<td>11.4</td>
<td>11.4</td>
<td>11.4</td>
<td>10</td>
<td>5</td>
<td>11.4</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>2.4</td>
<td>4.5</td>
<td>4.2</td>
<td>3.8</td>
<td>7.7</td>
<td>7.11</td>
</tr>
<tr>
<td>Teenage well-being</td>
<td>12.8</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>12.7</td>
<td>12</td>
<td>12.4</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12.8</td>
<td>12.2</td>
<td>12.6</td>
<td>12.48</td>
</tr>
<tr>
<td>Wonderful world of inventions</td>
<td>13.4</td>
<td>13.2</td>
<td>13.2</td>
<td>12.8</td>
<td>13.2</td>
<td>13</td>
<td>11.6</td>
<td>13</td>
<td>12.5</td>
<td>12</td>
<td>13.3</td>
<td>13.6</td>
<td>12.6</td>
<td>8.6</td>
<td>12.58</td>
</tr>
<tr>
<td>Traveling</td>
<td>10.1</td>
<td>8.9</td>
<td>4</td>
<td>4.1</td>
<td>10.6</td>
<td>8.9</td>
<td>6.8</td>
<td>11.3</td>
<td>7.1</td>
<td>2.1</td>
<td>3.7</td>
<td>3.6</td>
<td>4.1</td>
<td>0.7</td>
<td>6.58</td>
</tr>
</tbody>
</table>

Figure 6: Graphic representation of Granić et al.’s (2009) evaluation results

Based on the data collected and subsequent analysis, Granić et al. (2009) arrived at descriptive conclusions about the evaluated scenarios (for example, the learning designs
were based on constructivist and cognitive foundations; the Social Sciences and Student Research Project scenario grouping followed most closely the overall trend; the grouping of ICT scenarios that is least representative of the overall trend of pedagogical dimensions. The results do not indicate or lead to any evaluation judgement that distinguishes comparatively superior learning designs from comparatively inferior instances.

Therefore, based on the studies by Quinton et al., (2005) and Granić et al. (2009), it can be concluded that to utilise Reeves’s (1994) model, the number of dimensions changed for different evaluation purposes and use the refined model to derive a deeper understanding of eLearning environment designs in terms of their pedagogical foundations. However, in the context of the TELT Evaluation Framework, Reeves’s (1994) model does not help to judge which technology-enabled learning environments are more effective unless additional evaluation criteria are assigned to each pedagogical dimension so that the ratings provide more meaningful evaluation implications.

It is further suggested that Ho et al.’s, (2009) quantitative approach could be incorporated with Reeves’s (1994) model as their methodology can create criteria base on stakeholders’ requirements and thus serves to make valid evaluation judgements. Alternatively, it is possible to construct a reference system that synthesises Reeves’s (1994) model, Reeves and Hedberg’s (2008) evaluation heuristics, and/or similar heuristics developed by other researchers. For example, one of Reeves and Hedberg’s (2008) heuristics is “If the learner answers a question incorrectly, is he/she told the correct answer and why the answer given was wrong, if this is instructionally appropriate?”. This heuristic succinctly represents the “value of errors” dimension in Reeves’s (1994) model.

2 The Technical Layer (under development)

The technical layer of the TELT Evaluation Framework applies to the technical standards as specified and observed by UNSW, and the related issues and needs that require resolution in relation to the technical attributes of the TELT Platform. Where the technical standards are concerned, there is a need to identify the industry standards that are common to the eLearning software applications approved for the TELT Platform as applied to the various categories of applications outlined in this document. These categories will change and expand as new technical standards, requirements and applications emerge. Therefore, the technical requirements for TELT Platform applications must also be subject an iterative evaluation process.

In essence, the Stage 3 Technical Layer explores information technology within the University context. Amongst other functions, it:

i. Identifies technical eLearning standards and industry platforms
ii. Assesses the technical needs, capabilities, strengths and weaknesses of institutional users and IT infrastructure
iii. Informs the evaluation of any technical eLearning integration potential issues and pitfalls, and also information technology best practices to deliver successful platforms conducive to effective learning and teaching
Thus, the Technical Layer aims to deliver a framework for on-going assessments of eLearning technical compliance, feasibility, and effective integration with the whole community of users and institutional information technology.

Full compliance with technical standards and ongoing resolution of the issues and needs that arise in the use of the application features ultimately serves to ensure seamless interoperation, integration, adaptation, compatibility, and security. Moreover, rigorous observance of the technical compliance process also ensures thorough insight and familiarisation of each product's features may be gained as a result of using and testing to determine their strengths and weaknesses.

As for the technical issues, there is a need to identify the problems that commonly occur in the use of different types of eLearning products and to evaluate how these products are designed to minimise any negative impacts. Technical issues are not just about the applications. They also relate to any problems that are associated with the implementation process that includes for example, the user's required technical knowledge, the internal technical support that is needed, and considerations of users that require access to accessibility options.

(as a start) references to be considered in this section:

(Ismail, Idrus, Ziden, and Fook, 2009)

(Curriculum_Corporation and Education.au_Limited, 2002)

(Khedo, 2005)

(Franch and Carvallo, 2003)

(Graf and List, 2005)

(Sicilia and García, 2003)

(Fichten, et al., 2009)

(Torres, Juárez, Dodero, and Aedo, 2009)

(Saade and Kira)

3 The Business Layer (under development)

The business layer is primarily concerned with the range of linked activities that create benefit and value for users and the institution through processes and strategies that transform learning and activities (input) into more effective and valuable outcomes (output)
in the context of deploying eLearning applications. Constant technological change and refinements to learning and teaching approaches will always mandate change in business processes. Thus, there is a need not only to evaluate TELT applications in line with the agreed business processes identified for the TELT Platform, but also to subject the business processes per se to an iterative evaluation strategy to determine whether further refinement is required.

The Stage 3 Business Layer seeks to create value for users and the institution by integrating teaching and learning innovation with organizational structures and procedures, and reviewing business processes and operations, to create valuable and effective strategically aligned TELT outcomes. Amongst other functions, it:

i. Identifies formal business processes and procedural requirements, standards and practices for organizational change across both the University-wide and Faculty specific contexts

ii. Focuses on and assesses the business and organizational policies, guidelines, reports, staff and their responsibilities, procedures of relevant services and operation, planning of future operation, quality control, training and support, and vendor relationships and management as relevant to their own iterative evaluation and refinement, and their relationship with the TELT Evaluation Framework and teaching and learning practices

iii. Informs the Evaluation of any potential business processes and organizational change procedures, issues and pitfalls, and operational best practices to deliver successful, and institutionally compliant, teaching and learning improvements and innovative outcomes.

Thus, the Business Layer aims to deliver a framework for on-going assessment of and compliance with business and procedural requirements, operational feasibility, and the effective relationship between eLearning implementations and organizational standards and institutional practices.

In general, the TELT Platform business processes evaluation cycle will need to focus on: policies, guidelines, reports, staff and their responsibilities, procedures of relevant services and operation (for example, staff and student training, support, educational design), planning of future operation, quality control, vendor relationship and management.

(as a start) references to be considered in this section:


(Hughes and Attwell, 2006)

(Bull, 1999)

(Snabe, Rosenberg, Møller, and Scavillo, 2008)

(Balasubramanian and Gupta, 2005)

(Filipowska, Kaczmarek, Kowalkiewicz, Zhou, and Born, 2009)
The Benvic Benchmarking Project by Sangrà et al. (2001) focused on the establishment of evaluation criteria in order to achieve Quality Standards for Virtual Campuses. They analysed selected case studies and associated validation processes and subsequently released a series of indicators and competencies for benchmarking. Eight core meta-indicators are used in the Benvic System, namely:

- learner services
- learning delivery
- learning development
- teaching capability
- evaluation
- accessibility
- technical capability
- institutional capability.

Each of the eight indicators is associated with a range of assessment measurements that enables BENVIC users to carry out initial benchmarking diagnostics. The assessment measurements comprise three types:

- structural measurements
- practice measurements
- performance measurements.

The following is one example of the eight checklists that were devised and based on the eight indicators and the three types of measurement. The remaining seven checklists are fully available in the published document:

- **Structural Measurements**
  - Availability of systems and services to support communication among students and staff (inter-intra faculty)
  - Administrative procedures for enrolment of students are operational.
  - Cost-monitoring tools for student services are implemented.
  - Training services and materials for the staff in charge of learner’s services are available
  - Software to collect statistical data on student and staff utilisation of support services is implemented.
  - Technological solutions able to provide individual and group support are available.
  - Technology based measures to monitor and rectify common technical problems are adopted.
  - Support staff regularly briefed on the most frequent questions and problems encountered by students involving services.

- **Practice Measurement**
  - Measures to avoid dropping out after the guidance phase are adopted.
  - Throughout the course students are provided with technical and pedagogical support for using the services available (by means of written materials face to face sessions and online help).
• Performance Measurement
  • % of total staff trained in delivering the guidance process
  • % of applicants who undertook a guidance process
  • % applicants successfully enrolled
  • student support costs as % student course costs
  • student/technical support staff ratio
  • no. of student uses of support services per annum

4 Conclusion (to be completed once main document is established)
References


