EDITH COWAN UNIVERSITY
PERTH WESTERN AUSTRALIA

PUBLICATION - CMM5102

EvaluRater - A courseware evaluation tool that assesses both content and multimedia delivery for optimum instructional outcomes

by Serge Walberg

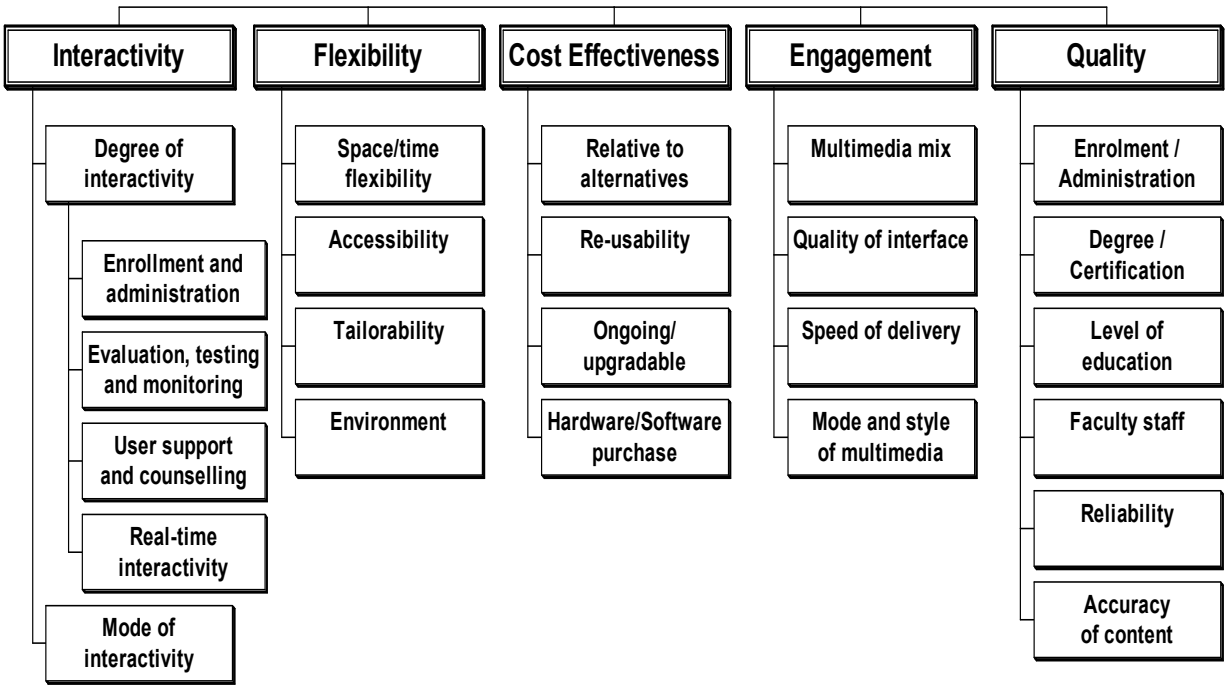
> "*Summative evaluation is evaluation done after software design and production is complete in order to establish its performance and properties.*" Steven Draper (1997)

With the panacea of on-line courses becoming available every day, there is a critical need for some reliable method of evaluation. The inexperienced user is unlikely, and often incapable, of examining each available alternative, or of investing much time and effort into assessing them.

EvaluRater was developed to enable both learners and on-line educators to conduct a comparative evaluation of the on-line courseware available on the market. It is completely customisable by the user, rendering it more effective in providing accurate evaluations specifically focused on the user's priorities. The tool is described below, and is accompanied by extensive instructions on its correct use. of the tool. This 'User Manual' explains not only the general use of the tool but also the techniques for customising and recycling it.

A question of primary concern is that of the *relative effectiveness* of on-line learning and traditional, *face-to-face* learning. For on-line learning to be preferable it must fulfill a set of vital requirements that address the following questions:

- Does it provide as much interaction between teacher and learner? (interactivity)
- Is it convenient as far as times and place? (flexibility)
- Is it more expensive? (cost-effectiveness)
- Is it less engaging or stimulating? (engagement)
- Is the quality of the education received as high? (quality)



Many other factors must be considered of course, but for our purposes we shall assume that most of them are somehow related to these 5 prime considerations. The chart above illustrates the different perspectives from which each one of these main considerations can be examined.

Before describing the tool and its operation, it is important to explain what each one of the evaluation criteria means, and how important that factor is in comparison with the other 4 factors.

1. Interactivity

This is the degree of communication and feedback between the learner and the teacher. It exists in a programme

- where the learner can ask questions and receive answers from the teacher,
- where work submitted by the learner is corrected and returned by the teacher,
- where the learner's details are recorded and registered,
- where the learner can access a variety of information packages and media by navigating through the programme
- where the programme will respond in different ways (reactively) depending on the learner's input. In this case the programme itself acts as a "tutor", assessing and correcting the user's input.
- Where a user support system is in place for learner guidance and counseling

These factors make interactivity a very important criterion, and one that should be given more consideration than all the others when evaluating multimedia. Barker & King[1], who produced a methodology for evaluating multimedia in 1992, state: "*Interactivity is an important hallmark of quality. Products that actively involve users in participatory tasks rate more highly*."[2]

Barker & King were primarily concerned with quality of *media* and *interfaces* rather than the effectiveness of *educational tools*, so their work is not specifically relevant here, but we shall borrow from them what is relevant to our purpose.

Most researchers agree on the importance of interactivity. Morrison[3] considers it *the* major criterion to be used in evaluation, and describes it as: ".. *the learner in conversation with himself over the material to be learned*"

2. Flexibility

This is a measure of the degree of adaptability the programme has with regard to accommodating to the schedule, location and disposition of the learner. A flexible delivery system allows the learner to learn at *hir* [4] most convenient time and place.

- Learner can work at home or other convenient location, with minimum travel requirements
- Learner can work at any convenient time, and is unrestricted by broadcast schedules, meetings, seminar or lecture times.
- Learner can work at *hir* own pace, without being delayed by slower-learning peers
- Learner can repeat exercises, re-study lectures, or re-do tutorials as many times as is required

---

Barker and King's methodology identifies 12 basic categories for evaluating courseware. These are: 1.Engagement, 2.Interactivity, 3.Tailorability, 4.Appropriateness of multimedia mix, 5.Mode and style of interaction, 6.Quality of interaction, 7.Quality of end user interface, 8.Learning styles, 9.Monitoring and assessment techniques, 10.Built-in intelligence, 11.Adequacy of ancillary learning support tools, 12.Suitability for single user/group/distributed use.

[2] Barker P.G., & King T.R., (1993). Evaluating Interactive Multimedia Courseware-A Methodology. Computer Education. Vol. 21, No 4, pp. 307-319

[3] Morrison D. J., Interactive (1987). Learning Systems and the Learner. Aspects of Educational Technology. Vol. XX, pp. 134-138

[4] Politically correct, gender non-specific form of his/her.

- Learner can *customise* the programme to *hir* preferences (style, mode, *pace*, media, etc.)
- Programme is adaptable to various platforms, operating systems, networks, etc.

Flexibility is another pivotal evaluation criterion, and deserves close attention. It is certainly the single most persuasive argument in favour of distance education, as opposed to *face-to-face* learning.

3. Cost-effectiveness

Today, with education becoming just one more marketable consumer commodity, cost and value are becoming critical factors in any evaluation of learning programmes. Corporate entities requiring an ongoing training programme for their human resources, for example, will be primarily concerned with the *cost-effectiveness* of any programme adopted.

A cost-effective on-line learning package

- Is more economical than a face-to-face programme (requiring the teacher, the learner, or both to travel).
- Can be re-used, resold, recycled or replicated so that the cost is amortised by several populations of learners.
- Is easily, and cheaply 'upgradeable', thereby avoiding the need to repurchase a new product when content is out-dated.
- Does not require investing funds into new or expensive hardware, software, or human resources.
- Does not require expensive or time-intensive pre-training of learners in the *use* of the programme.

While cost-effectiveness may not seem as important a factor as flexibility for an individual learner, it would be a major consideration for, say, a multinational corporation with tens of thousands of employees to re-train.

4. Engagement

At the end of the day, the only thing that guarantees the programme is actually going to work is if the end-user enjoys it enough, and feels stimulated enough to pursue the training. In some cases (such as programmes for toddlers for example), this feature can be far more important than others (such as accuracy, cost-effectiveness or interactivity)

Much has been made in the literature of the 'engagement' factor, some researchers placing it at the top of the list of determining variables. Laurel [5](1990) stresses: "Everything about the interface should *engage* the user to accomplish the task".

Reeves[6], who devised a rating tool called *Interface Dimensions* in 1993, lists 10 "dimensions" which his tool grades on a linear scale from *Easy* through *Difficult*. An engaging interface, he says

- Is easy to use
- Is easy to navigate
- Puts little pressure on learner to remember procedures (*cognitive load*)
- Maps user's path through the programme
- Has well designed screens
- Presents information well
- Integrates media well into the programme
- Has an attractive and appealing "look" and "feel" to it (*aesthetics*)

To Reeves' Rating Tool we shall add one more factor, with which he was not concerned (since his focus was primarily on interface design):

[5] Laurel, B., (1990). The Art of Human-Computer Interface Design. Reading, MA: Addison-Wesley
[6] Reeves, T.C., (1993). Systematic Evaluation Procedures for Instructional Hypermedia/Multimedia. American Educational Research Association Journal.

- Speed of delivery. The learner can undergo the training in a reasonable time-span, (no long delays in receiving course material or teacher feedback)

With the spiraling development of new technologies in 2D and 3D graphics, Virtual Reality, audio and video multimedia, a wide spectrum of embellishments, add-ons, animations, etc. are becoming available to interface designers. This means that learning programmes can provide increasingly more engaging, user-friendly interfaces.

5. Quality

Last but certainly not least is the quality of the education, training and final degree or certification awarded. In some instances, for example in the "flight-simulator" training of airline pilots, the quality of the training and certification received far outweighs any other consideration.

The quality of the faculty and teaching staff is primordial. On-line teaching requires skills and capabilities very different from those expected of a classroom teacher. Where a *systems approach* is in place, the quality of the delivered product is as good as the weakest link in the systems chain; "Are all teachers capable of becoming proficient and effective with any medium?" ask Moore and Kearsley[7].

The quality of the on-line or web-based learning programmes can be determined by assessing the following factors:

- The existence of enrollment and registration procedures, with the learner's details recorded.
- The quality and *bona fide* value of the certification awarded at the end of the training.
- The quality of the actual teaching delivered in terms of content, instructional design, etc.
- The teaching skills of the programme coordinators.
- The reliability of the course and it's parent institution, it's reputation, history, credentials, etc.
- The accuracy of the content of the delivered material and instruction

With education fast becoming a *user-pay,* profit-motivated, training scheme for corporate business' human resource requirements, a real danger exists that the quality of education delivered will deteriorate in the interest of economic rationalism. It must therefore be a primary concern for educators and learners alike to impose, demand and maintain as high a quality as possible on any learning programme.

EVALURATER - AN ON-LINE COURSEWARE RATING TOOL

I designed and constructed EvaluRater to take into account not just the degree to which each criterion applies to a particular product, but also the relative importance of each criterion. This proportional "weighting" of the tool makes it uniquely customisable and applicable to most types of educational software.

Each of the five main criteria described above is composed of a series of sub-criteria. To evaluate a product, each item is assessed against the product and rated 0, 1, or 2. For example: assessing **Interactivity** under the item "enrollment and registration",

- If none exists a score of **0** is given.
- If some exists :**1**
- if a very efficient and effective procedure of enrollment and registration exists, the item would rate **2**

When all the *sub-categories* in one main category have been rated, their score is totaled. The value returned is then multiplied by a "Weighting Index" specifically determined for that category. The

[7] Moore, G.M., & Kearsley, G., (1996). Distance Education. Belmont: Wadsworth Publishing company.

Weighting Index shall be less than **1** (i.e. from **0** to **0.9**) and makes the instrument *completely customisable.*

After rating has been applied to all 5 main categories, the grand total is calculated, and the value returned is used for comparative evaluations of different programmes.

The critical importance of the Weighting Index is due to its capacity for *transfer and application*, or it's ability to be customised or adapted according to the user's priorities. These allocations of values must add up to 1.0 in order for the tool to function correctly. For example, if the courseware was targeting *very young children*, the Weighting Factors might be set as follows:

INTERACTIVITY:	**0.2**
FLEXIBILITY:	**0.1**
COST EFFECTIVENES:	0.**0**
ENGAGEMENT:	**0.4** (most important factor)
QUALITY:	**0.3**

If, however, it was training corporate secretaries of a multinational company:

INTERACTIVITY:	**0.1**
FLEXIBILITY:	**0.2**
COST EFFECTIVENES:	**0.5** (most important factor)
ENGAGEMENT:	**0.1**
QUALITY:	**0.1**

The following table shows an example of a template for using the tool

INTERACTIVITY		**FLEXIBILITY**		**COST-EFFECTIVENESS**		**ENGAGEMENT**		**QUALITY**	
Admin/enroll		Location		Alternatives		Multimedia		Certification	
Test/monitor		Time		Re-usable		Interface		Level	
Support		Access		Upgradable		Speed		Faculty/staff	
Real-time		Tailorability		Purchases		Aesthetics		Reliability	
Other		Other		Other		Other		Accuracy	
								Other	
TOTAL		**TOTAL**		**TOTAL**		**TOTAL**		**TOTAL**	
Multiply by		Multiply by		Multiply by		Multiply by		Multiply by	
Weighting Index		WeightingIndex		Weighting Index		Weighting Index		Weighting Index	

Total a + b + c + d + e

a + b + c + d + e = **Assigned Evaluation Rating**.

Blanks have been intentionally left in the template so that the user can *further* customise the tool by adding other criteria important to *hir*.

Care must be exercised not to expect perfect evaluations of entire learning systems by simply interpreting a series of numbers, which at best can give only a comparative evaluation. In order for the tool to deliver more accurate evaluations, a textual assessment can be entered by the user in response to each sub-category item.

The tool can therefore be used in 2 distinct ways:

- to give a broad comparative evaluation using a simple arithmetic calculation
- to provide an in-depth written evaluation using the tool criteria template.

DEMONSTRATION OF EVALURATER ON SAMPLE SOFTWARE

Two on-line courses have been selected for evaluation in order to demonstrate the use of EvaluRater, these are:

1. **8 Minute HTML** @: http://drott.cis.drexel.edu/8Minute.html
2. **Learn Spanish** @: http://www.studyspanish.com/tutorial.htm

These 2 were chosen because they provide "good" and "bad" examples of on-lIne courses. **8Minute Html** is a relatively unsophisticated course, basically providing a linear on-line presentation of text, with little Interactivity, Quality or Engagement value. Learning Spanish is a more professionally designed teaching programme, fulfilling many of the criteria of an effective on-line course.

1. 8 Minute Html

For the rating tool to deliver an evaluation we must first customise the *Weighting Indices* for the 5 main categories; we shall assign the relative values we estimate to represent our priorities

INTERACTIVITY: 0.2 FLEXIBILITY: 0.3 COST-EFFECTIVENESS: 0.0 ENGAGEMENT: 0.2 QUALITY: 0.3

Next, we must assign values to all the sub-criteria using the template:

INTERACTIVITY		**FLEXIBILITY**		**COST-EFFECTIVENESS**		**ENGAGEMENT**		**QUALITY**	
Admin/enroll	0	Location	2	Alternatives	2	Multimedia	0	Certification	0
Test/monitor	1	Time	2	Re-usable	2	Interface	1	Level	1
Support	0	Access	1	Upgradable	1	Speed	2	Faculty/staff	1
Real-time	0	Tailorability	0	Purchases	0	Aesthetics	1	Reliability	1
Other	1	Other		Other		Other		Accuracy	2
								Other	
TOTAL	**2**	**TOTAL**	**4**	**TOTAL**	**5**	**TOTAL**	**4**	**TOTAL**	**5**
Multiply	**.2**	Multiply by		Multiply by		Multiply by		Multiply by	
Weighting Index		WeightingIndex		Weighting Index		Weighting Index		Weighting Index	

Totals = 0.4 + 1.2 + 0 + .8 + 1.5

Assigned Evaluation Value: 3.9

The Rating Tool can also be used to deliver a textual evaluation, by addressing each sub-criteria individually. A sample *textua*l evaluation of the product would render:

INTERACTIVITY:
- **Enrollment/administration:** none
- **Evaluation/testing:** some testing of learner codes
- **User support:** none
- **Real-time interactivity:** none
- **Other:** some useful links

FLEXIBILITY:
- **Location:** learner's convenience
- **Time:** learner's convenience
- **Accessibility:** all platforms, no special software
- **Tailorability:** none

COST-EFFECTIVENESS:

Relative to alternatives: much cheaper than normal course
Re-usability: can be re-used for many learners
Up-gradeable: not at all
Purchases: none needed

ENGAGEMENT:
Multimedia mix: none at all
Interface: bland but comfortable
Speed of delivery: instantaneous
Mode and style of media: not even graphics

QUALITY:
Certification: none
Level: medium
Faculty/staff: good
Reliability: average
Accuracy: good

2. *2. Learning Spanish*

INTERACTIVITY: 0.3 FLEXIBILITY: 0.2 COST-EFFECTIVENESS: 0.1 ENGAGEMENT: 0.2 QUALITY: 0.2

INTERACTIVITY		FLEXIBILITY		COST-EFFECTIVENESS		ENGAGEMENT		QUALITY	
Admin/enroll	1	Location	2	Alternatives	2	Multimedia	1	Certification	0
Test/monitor	2	Time	2	Re-usable	2	Interface	1	Level	2
Support	1	Access	1	Upgradable	1	Speed	2	Faculty/staff	1
Real-time	0	Tailorability	1	Purchases	2	Aesthetics	2	Reliability	1
Other	1	Other	1	Other		Other	1	Accuracy	2
								Other	
TOTAL Multiply Weighting Index	**4**	**TOTAL** Multiply by WeightingIndex	**7**	**TOTAL** Multiply by Weighting Index	**7**	**TOTAL** Multiply by Weighting Index	**7**	**TOTAL** Multiply by Weighting Index	**6**

Totals = 1.2 + 1.4 + 0 .7 + 1.4 + 1.2

Assigned Evaluation Value: 5.9

A textual evaluation would render:

INTERACTIVITY:
Enrollment/administration: Registration of learner
Evaluation/testing: On-going evaluation/ correction
User support: adequate help / support
Real-time interactivity: none
Other:

FLEXIBILITY:
Location: learner's convenience
Time: learner's convenience
Accessibility: all platforms, no special software
Tailorability: a small degree

COST-EFFECTIVENESS:

Relative to alternatives: much cheaper
Re-usability: can be used for many learners
Up-gradeable: not at all
Purchases: none needed

ENGAGEMENT:
Multimedia mix: medium
Interface: good
Speed of delivery: instantaneous
Mode and style of media: medium

QUALITY:
Certification: none
Level: medium
Faculty/staff: good
Reliability: medium
Accuracy: good

As anticipated, the Assigned Values delivered by the Rating Tool indicate that the second product is substantially more effective than the first one. This shows that our tool functions correctly, and that it can be customised to adjust to the priorities of an individual user. Furthermore, it is extendable/upgradeable, since the user can add *hi*r own criteria to the tool.

The Rating Tool described above will effectively and efficiently *assis*t in the evaluation of on-line learning programmes, primarily because of its customisability and flexibility. The tool's salient features are:

1. it addresses all the major criteria relating to courseware evaluation
2. it breaks down major categories into related sub-categories and permits a value assessment of each.
3. It allows the inexperienced learner to apply a simple rating scheme to each sub-categories
4. It is customisable, and can be adapted to different learners' priorities by adjusting the Weighting Indices

No rating tool can possibly evaluate any product with 100% accuracy, especially an instrument relying on simple arithmetic computations. Every on-line or web-based learning programme is unique, and generalisations can sometimes lead to erroneous conclusions.

Nevertheless, every attempt was made to build into the design of **EvaluRater** as many variables and relevant criteria as was considered necessary, to provide effective evaluations of on-line and web-based learning programmes.