# CRITERIA FOR EVALUATING INTERNET TUTORIALS IN SPEECH COMMUNICATION SCIENCES

*Chris Bowerman[1], Anders Eriksson[2], Mark Huckvale[3], Mike Rosner[4], Mark Tatham[5], Maria Wolters[6]*
[1]University of Sunderland, [2]Umea University, [3]University College London, [4]University of Malta, [5]University of Essex, [6]University of Bonn
mark@phonetics.ucl.ac.uk
http://tn-speech.essex.ac.uk/tn-speech

## ABSTRACT

The Computer Aided Learning (CAL) working group of the SOCRATES thematic network in Speech Communication Science have studied how the Internet is being used and could be used for the provision of self-study materials for education. In this paper we follow up previous recommendations for the design of Internet tutorials with recommendations for their evaluation. The paper proposes that evaluation should be seen as a necessary quality assurance mechanism operating within the life-cycle of CAL materials development. We propose a structured set of criteria for evaluation, based on the features of good tutorials, against which a tutorial might be judged. Since evaluation against fixed criteria is only one possible approach, we outline how evaluation could also be performed by student users and in controlled trials

## 1. INTRODUCTION

The Computer-Aided-Learning working group of the SOCRATES Thematic Network in Speech Communication are studying the current use and the future potential of the Internet to support teaching and learning in the field [3,5]. Our findings are that Speech Communication education has special characteristics that lead to particular interest in Internet/CAL: it is strongly interdisciplinary, distributed over many types of university departments, involves multiple languages and usefully exploits interactive media.

However our survey of existing Internet teaching resources in the field [4] highlighted the fact that much educational material on the Web was poorly designed from the students point of view. We found material that was no more than a text-book presentation of ideas, material that did not encourage the student to discover new concepts, material that did not challenge the understanding of the student, and material that did not make the best use of the medium.

In [3] we proposed that progress could be made through the building of small autonomous teaching components that we called, simply, *tutorials*. The key features of tutorials were that they (i) were oriented to self-study, (ii) had restricted scope and pre-requisites, (iii) had restricted conceptual difficulty, (iv) had a defined internal structure, (v) were largely self-contained, and (vi) contained mechanisms for self assessment. Our aim was to define a re-usable teaching component that was small enough to be written by one person.

In [1,5] we discussed general principles of courseware development and made recommendations that should be useful to any current or potential author of tutorial materials designed for use on the Web. In this paper we turn to the topic of the evaluation of tutorials. The evaluation of CAL is a complex area concerning fitness for purpose and cost-effectiveness as well as meeting learning needs. In section 2 we propose that evaluation should be seen as part of the life-cycle of CAL development. In section 3 we propose some general features of good tutorials against which a tutorial might be judged. In section 4 we emphasise that evaluation against fixed criteria is itself only one possible means of evaluation and discuss how other measures of effectiveness might be obtained.

## 2. ROLE OF EVALUATION IN CAL

### 2.1 Lifecycle

A CAL package is both a software artefact and a piece of courseware. From its original conception to its production and in-service maintenance, the team working on CAL courseware move through the lifecycles of both courseware design and software production.

In outline, both these lifecycles involve a phase of the analysis of the user's needs and the aims and scope of the system to be built, followed by a phase in which the system is developed, tested and evaluated, before going into service. (see, e.g., [2,6]) The key point that we would wish to stress is that evaluation is a natural part of the development of CAL software. We shall look at why?, what?, how? and when? CAL might be evaluated.

There are two main reasons for engaging in evaluation: either for a course developer seeking *formative* information on how a course might be improved or for a course tutor wishing to gain *summative* evidence of areas in which the existing, running course may be improved for future cohorts of students.

The areas in which evaluation may take place would, for developers, include the need for the course, whether or not a market for it existed and whether or not it provided a sufficient payoff. The production of one hour of multimedia CAL could take half a man-year of development, therefore unless the use of the CAL medium offers particular advantages (less risk, less costly, more repeatable, increases access), it will not necessarily prove to be a viable product. For end-users of developed courses and evaluators of material under development, the course must be compared to other existing systems and courses in order to determine whether or not the new course achieves the desired learning outcomes more effectively and without any unwanted side-effects. Typically evaluators would look for greater problem solving and skills mastery, improved information retention and evidence that any desired attitudinal changes having taken place. CAL may be evaluated at different time periods: either pre- and post-course, or periodically during the course. It is often important not merely to concentrate on the end outcomes of a course but the process of learning itself – this can reveal skills changes over the duration of the course which are less easily measured by traditional exit tests.

We have motivated the need for evaluation in terms of being a phase of CAL production in order to improve courses under development and inform course tutors on how best to run existing courses.

## 2.2 Evaluation perspective

CAL may be evaluated from different perspectives. These relate to the different needs of the stakeholders in a CAL system: developers, tutors and students.

The software developers wish to ensure the robustness of the system. This will not only include the verification and validation of the system but also its computational efficiency and the ergonomics of the user interface.

Tutors using the system wish to assure themselves of its pedagogic viability. This will include a consideration of the content and the teaching and learning approach taken. Tutors might also compare the effectiveness of the CAL approach to their previous teaching method, in order to ensure that they employ the most effective tools in each teaching situation.

It is also worth contemplating evaluation from the student perspective. Students wish to ensure that the system meets their learning needs and is straightforward to use.

## 3. EVALUATION CRITERIA

In this section we list some criteria against which developers and experts might evaluate tutorial material. Such criteria come from observations of what is good practice in the field.

## 3.1 Goals and Objectives

### Definitions

Goal: a general description of the intended outcome

Objective: a specific description of details of the intended outcome

Content: The material that is used in the tutorial.

A goal is achieved by objectives. An objective is realised by content.

### Preliminary Evaluation Questions

- Are the goals and objectives clearly enough stated to
  ⇒ enable re-use,
  ⇒ enable formal evaluation,
  ⇒ match objectives to goals?
- Does achieving the objectives realise the goals?
- Are the target audience and pre-requisites clearly stated?

Evaluation of teaching materials only makes sense in the context of a clear statement of their audience. It is here that we can begin to introduce evaluative notions such as reasonableness and appropriateness in order to give meaning to the term 'effectiveness'. Materials are only effective in terms of the author's intentions, and audience is an important parameter here.

## 3.2 Content

### Coherence, consistency and correctness

- Do the materials form a coherent whole?
- Are the materials free from contradiction and is the use of technical vocabulary consistent?
- Is the subject material accurate and up-to-date?

### Match between Objectives and Content

- Does the content appropriately reflect the stated objectives?

### Reasonableness and Appropriateness

- Is the content *reasonable* given the intended audience with respect to
  ⇒ The required skills and background knowledge,
  ⇒ The allocated time,
  ⇒ The available resources.
- Is the content *appropriate* with respect to the overall programme of study?

## 3.3 Teaching and Learning

### Cost Effectiveness

- Are the materials designed in such a way that it is possible to measure their cost effectiveness in terms of learning pay-off?

- Do the materials deal with subject matter which is best taught using this method?

### Open Learning, Assessment and Feedback

- Are the materials fully self-instructional?
- Are there means for self-assessment that match the target skills and knowledge defined by the objectives?
- Are there means for feedback to both student and teachers?
- Are there means for the student to communicate with a teacher if there are difficulties?
- Does the tutorial create and maintain learner motivation and interest?
- Does the tutorial encourage active learning?

## 3.4 Implementation

### Topology

- Is the topology the best way of implementing the course content and its objectives?

### Re-usability

- Is the design of the material suitable for re-use by other authors?
- Is the design and content presentation suitable for translation to other languages?
- Are the computational requirements clearly stated?
- Is the tutorial easily portable to other platforms?

### Computer Literacy

- Are the materials designed in such a way that their use is clear from the computer literacy point of view?
- If complex operations are involved is there an appropriate optional element of computer literacy skill development?

### Balance, Flexibility and Flow

- Are subsections of the material allocated the appropriate time and content?
- Can parts of the module be skipped?
- Can users incorporate their own material at certain points (i.e. is it possible branch off to some other material)?
- Is there a logical progression between the points being made?
- Do later points reinforce the earlier points?
- Is the logical progression obvious to the students? If something is not obvious to the students (for a particular reason) is it obvious to the teacher?

## 4. EVALUATION IN PRACTICE

### 4.1 Evaluation during the Software Lifecycle

Creating a tutorial is best seen as an evolutionary process: an initial requirements analysis is followed by several iterations of design, implementation, and evaluation. In this approach, the designer can start with relatively straightforward evaluation techniques and then move on to more elaborate methods the more functional the resulting prototypes become.

But evaluation should be more than an independent stage in the development process. The checklist we provide in Sec. 3 allows designers to judge how their choices will affect the quality of the final tutorial during the stages of requirements analysis (goals and objectives, 3.1), design (3.2–3.3), and implementation (3.4). It is also a good basis for deciding on which aspects need further advice from experts and potential users.

### 4.2 Assessing the Effect on Student Performance

A tutorial has had the desired effect if it helps students achieve the learning objectives more easily. This can be determined in a variety of ways. A good first step is to observe students using the material and protocol their problems with form and content. This can also reveal if, after working through the tutorial, students are able to perform certain tasks more quickly and confidently, such as deciphering mystery spectrograms. References [8,URL1] are good sources of ideas and techniques, while [7] provides a more general view of course evaluation.

It is also possible to monitor whether student scores are improved or the dropout rate has decreased in courses where the tutorial was used. Here it may be hard to establish a causal link between the changes and the use of the tutorial, and this is only feasible at a very advanced stage of development.

### 4.3 Getting Student Feedback

Feedback from students can be collected in interviews, either in person or through a (student) intermediary, or by (anonymous) questionnaires. For more stable versions, a group of beta testers can be established which are instructed to report comments or problems back to the design team.

The advantage of questionnaires is that they can be used to collect feedback from a large number of users without further contact. However, questionnaires are rather complex measurement devices which need to be carefully calibrated to become both reliable and interpretable. They are also notoriously sensitive to the circumstances in which they are administered. Therefore, they should be complemented by other evaluation activities, such as interviews or observation.

The questionnaire items should cover all aspects of the course, from contents to delivery, from navigation to problems with network connections, and especially on those aspects you would like diagnostic feedback. Closed answer forms with fixed ratings or Yes/No-boxes are easier to analyse than open questions; however, in open answers, students may mention problems you had not thought of previously. As a good compromise, you could conclude a set of rank-based evaluations by the questions "What did you like best/worst about this tutorial?"

A good strategy is to copy some questions from a validated course questionnaire and to add questions relevant to your specific objectives. If you ask users if they agree with certain statements about the tutorial on a scale of 1–5 or 1–7, it is helpful to change the polarity of the statements from positive to negative once in a while; this catches respondents who insert values mechanically. At [URL1], you can find a sample questionnaire for the evaluation of web-based tutorials. Finally, remember that long questionnaires with lot of open questions are less likely to be filled out and returned than short, concise ones and set your feedback priorities accordingly.

### 4.4 Conducting Formal Evaluations

Full formal evaluations are in effect large experiments. They need to be carefully designed taking into account the relevant pedagogical literature. Therefore, they should only be conducted once the tutorial is already stable and well-tested using less elaborate methods. However, proper evaluations are essential if the tutorial is to be distributed commercially.

For a formal evaluation, you first need to specify what you want to measure, e.g. acquisition of skills, changes in student attitude, or understanding of concepts, and how this connects to your original set of goals and objectives. The experiment design should include at least one control group, and there are many potential distortions which might affect results: students' intrinsic motivations, their previous experiences with the subject, the fact that new delivery methods are always more exciting than old ones, the attitude of the teacher of the course, etc. Lastly, the evaluation should be conducted in collaboration with an experienced educationalist.

## 5. CONCLUSIONS

In this paper we have tried to construct a procedure for the evaluation of Internet tutorials in Speech Sciences. In all our discussions, we are not of the opinion that self-study tutorial materials *alone* form an adequate means of tuition in our field. However good quality tutorials can serve as an adjunct to conventional teaching or as a partial replacement; and evaluation plays one part in assuring their quality.

For an author starting out on the construction of a tutorial we would recommend that early consideration is made for the dissemination of the teaching material. If a tutorial is not intended to be shared, we think it unlikely to be a cost-effective use of the author's time.

## 7. REFERENCES

[1] Bowerman, C., Eriksson, A., Huckvale, M., Rosner, M., Tatham, M., Wolters, M., "Tutorial Design for Web-based Teaching and Learning", ESCA/Socrates MATISSE Workshop, London, April 1999.

[2] Gagne, R., Briggs, L., Wagner, W. (1992). *Principles of Instructional Design*. Harcourt Brace Jovanovich.

[3] Huckvale, M., Benoit, C., Bowerman, C., Eriksson, A., Rosner, M., Tatham, M., Williams, B. (1997a). Computer Aided Learning and Use of the Internet. In *The Landscape of Future Education in Speech Communication Sciences: 1 Analysis*, G. Bloothooft *et al.* (eds.), Utrecht: OTS Publications.

[4] Huckvale, M., Benoit, C., Bowerman, C., Eriksson, A., Rosner, M., Williams, B., (1997b). Opportunities for Computer-Aided Instruction in Phonetics and Speech Communication Provided by the Internet. *Proceedings of the 5th Eurospeech Conference*, Rhodes.

[5] Huckvale, M., Bowerman, C., Eriksson, A., Rosner, M., Tatham, M., Williams, B., Wolters, M. (1998). Computer Aided Learning and Use of the Internet. In *The Landscape of Future Education in Speech Communication Sciences: 2 Proposals*, G. Bloothooft *et al.* (eds.), Utrecht: OTS Publications.

[6] Partridge, D. (1992). *Engineering Artificial Intelligence Software*. Intellect.

[7] Ramsden, P. (1992): *Learning to teach in higher education*. London: Routledge

[8] Ramsden, P. / Dodds, A. (1989): *Improving Teaching and Learning: A Guide to Evaluation*. Melbourne: Centre for the Study of Higher Education, University of Melbourne

[URL1] Evaluation of technology-based teaching and learning.
http://www2.deakin.edu.au/dcad/FTL/evaluation