

Is E-Learning Up to the Mark?

Fundamentals in Evaluating New and Innovative Learning Approaches Involving Information- and Communication-Technology

Oliver Haas¹

Abstract

With the introduction of training courses or seminars via the internet it becomes relevant to evaluate and assess these new and innovative forms of learning and teaching. The following article deals with two issues:

- a) Evaluation methods used in web-based learning arrangements that are dependent on standards in evaluation research
- b) The implicit logic of the communication medium “internet”.

The paper concludes by illustrating methods of evaluation and assessment that correspond to both paradigms.

¹ Oliver Haas (M. SocSc.) conducted university studies in Sociology at the Johann-Wolfgang v. Goethe University of Frankfurt/ Germany and the Free University of Berlin/ Germany. He is currently employed a Technical Advisor by the [German Agency for Technical Cooperation](#) (GTZ) and has worked in Russia, Tanzania, Malaysia, and South Africa. Here he has been involved in Vocational Education and Training projects.

Introduction

Does the internet make us more intelligent? Do we obtain more knowledge, skills and better qualification through its utilization compared to the “traditional” methods of teaching and learning? Does the internet change teaching and learning concepts or even our perception of learning? After the last big technical innovation—print media—had a significant impact on teaching and learning, the most recent learning technologies² only changed learning strategies in human resource development to a minor extent. However, the most recent development in Information and Communication Technology (ICT)—the world wide web—as a universal medium of exchanging and mediation of information and knowledge has set a new yardstick regarding the accessibility of learners and the dissemination of learning content.

Innovations in general, and innovative teaching and learning procedures specifically, are always under pressure to legitimize their existence and benefits. This becomes even more crucial when their application is highly related to risks and financial expenditure. Efficiency and effectiveness, quality as well as relevance and significance, to name a few, are key criteria under which these procedures are critically assessed. Despite the recent enormous interest generated in learning via the web³, the question of acceptance, efficacy and suitability has not been answered fully.

² E.g. language laboratories, Computer based training, etc.

³ The terms “web-based learning”, “online learning”, “internet-based learning”, and “e-learning” are used interchangeably in this paper.

Methodology-driven and comprehensible criteria-based assessments of procedures, events or actions are called “evaluations”. So far evaluations of internet-based learning have been mainly applied in the public sector. Yet over the past years, the question of efficacy and efficiency of online-learning has become increasingly relevant for the private sector. With regard to in-house and on-the-job training the common understanding is that any training is an investment in employees that needs to be justified like any other investment. While the determination of “costs” is relatively easy to define, the determination of “economies of scale” has proven to be a challenge on its own. This is where evaluation becomes relevant.

To evaluate ‘learning via the internet’ requires an accurate preparation of empirical research that can only lead to useful data if meaningful criteria under which the evaluation will take place have been assigned in advance. Up to now evaluation of learning via the web has been a seriously neglected aspect of impact assessment in education and training. As such, theories on how to go about evaluating learning via the net as well as gaining empirical valid data are difficult to find. However, evaluations in this field are not just necessary but also possible.

It is the aim of this paper to provide a selection of possibilities on how to evaluate the efficiency and benefits of online learning. Due to the complexity of e-learning, this is not possible without a methodological introduction.

Components in evaluating online seminars

To evaluate learning via the internet is nothing new. Assessments of learning software (e.g. in the field of Computer Based Training) have been conducted widely and provide useful experiences as well as theoretical concepts, paradigms and procedures. These are methodological aspects that cannot be neglected when evaluating learning via the internet.

In recent years, highly acclaimed work has been done in the development of methodical standards and instruments in evaluating the efficacy of education and training. However, these standards—consisting models, methods, and instruments—can also be utilized in other contexts. Evaluation methods and criteria are often explicitly created for a specific evaluation purpose. This suggests that evaluation—or better evaluation research—should be considered as applied science following a specific need and demand.

Areas and criteria of evaluation

The setting of criteria should be the first step when assessing online training courses. Criteria need to be defined for each and every aspect of the learning scenario. This involves all participants of the training course, the utilized learning material, the pedagogic approach, and technological aspects, as well as guidance in the learning process and technical-administrative support as part of the training course. Criteria are directly related to the quality of online seminars and constitute the foundation of any evaluative approach to online learning. Thus, a clear statement and definition of evaluation criteria is of crucial importance for the whole evaluation process.

Often several areas of evaluation are strongly interlinked and have a significant impact on each other.⁴ However, for a differentiated assessment of online learning courses it is essential to select evaluation criteria for each evaluation area separately. The decision on evaluation areas and its correlating evaluation criteria is to be done at the beginning of the overall evaluation and needs to be specified for each learning program. Yet, there are some criteria that could be defined as

⁴ For example, the influence of the pedagogic approach or the technology on the students' motivation as well as learning success.

typical in evaluating online learning. The following chart provides a selection of these criteria:

Chart 1: Evaluation areas und evaluation criteria for online learning

Evaluation area	Evaluation criteria
Participants/ students	<ul style="list-style-type: none"> • Acceptance of training course • Drop-out rate • Degree of collaborative learning⁵ • Rate and intensity of interaction with learning content • Learning success • Communication among students • Transfer and utilization of learning content at the workplace
Pedagogical approach	<ul style="list-style-type: none"> • Learning and teaching methods • “Didactic of activation”⁶ • “Didactic of enabling”⁷ • Degree of “blend” in the pedagogic approach⁸
Learning material	<ul style="list-style-type: none"> • Editing and processing of learning content • Comprehensibility, amplitude, correctness, “time sensitivity” of learning material

⁵ User of training courses can collaboratively work on tasks independent of time and space.

⁶ For a detailed explanation of the term please refer to page 4.

⁷ see above.

⁸ “Blended learning” is an integrated learning concept that combines Information and Communication Technology (ICT) with “traditional” learning methods and media in a single learning arrangement.

Technical system	<ul style="list-style-type: none">• Quality and reliability of connectivity• Technical infrastructure at the learning site (e.g. internet accessibility)• Collaboration and communication tools
Support and administration	<ul style="list-style-type: none">• Registration and financing• Online-support• Offline-support• Technical support

Participants

When evaluating online training courses it is crucial to remember that it is not the learner that is evaluated but the learning content delivery! However, the learner plays an important role as a resource person for evaluating the overall training program. The learner’s behavior, learning success as well as transfer of learning content to the workplace provide important empirical information with regards to the quality of the training course.

Pedagogic approach

Even though internet-based learning has been—especially in its early years—strongly related to technology, it is still about the provision of learning content and qualifying people in order to create employability. Therefore web-based learning also—or even more—has to be based on a pedagogic foundation, generally provided through the Curriculum. In Education and Training one distinguishes between two didactic models:

*“Didactic of activation”*⁹

Coming from engineering science, the “didactic of activation” assumes that successful learning can only take place if it is adequately planned with learning methods having been selected accordingly and the sequence of learning is being followed rigidly. Programmatic learning and curricular planning are core determinants of this model.¹⁰

*“Didactic of enabling”*¹¹

This model focuses on the learning and its success. It tries to create an enabling environment for the learner to build up on existing knowledge and to expand his skills and competencies according to his need and demand. Group work, the provision of several learning paths and methods to acquire knowledge are key features of the “didactic of enabling”.¹²

Degree of “blend” in the pedagogic approach

The “blended” learning approach treats web-based learning as one way of delivering learning content. Consequently, other methods of “traditional” learning

⁹ German: “*Erzeugungsdidaktik*”

¹⁰ See Arnold, Rolf/ Schüssler, Ingeborg, 1998

¹¹ German: “*Ermöglichungsdidaktik*”

¹² The “action-oriented learning approach” is one relevant learning approach of this model. The action-oriented learning approach is based on a holistic interpretation of technical, individual, methodological and social competence. Learners graduating through this approach are expected to have acquired not only skills and knowledge obtained from qualifications, but also “key competencies”, such as problem solving techniques, communication skills and the ability to work in teams (see Heitmann, Werner, 2004).

such as group or individual exercises remain valid and relevant. Depending on the Curriculum, it has to be decided which learning content will be delivered via the web and where other forms of learning material are more relevant. The demand for content delivery via the web must result from the pedagogic approach of the overall learning program.

Learning material

The learning material provided should be a main focus of the assessment in two respects:

- 1) Quality of content
- 2) The processing of learning content into learning material incorporating ICT

In terms of quality of content it is mainly comprehensibility, coverage, correctness and “time sensitivity” of learning material that needs to be evaluated. Concerning its processing of content into learning material the evaluation should focus on the conversion of content to learning material (e.g. the utilization of text, pictures, animation, simulation, etc.).

Technical System

When dealing with web-based learning, technical aspects are of great importance. Quality and reliability of connectivity including time needed for loading frames and content need to be assessed. The stability of the Web-server as well and its capacity are aspects that must also be assessed. Additionally, the extent and time of utilization of web-content and services provided by the learning platform can provide useful information on the suitability of the technology.

Support and administration

E-learning can take place on an independent level or as an add-on to teaching (the 'blended' approach). The content-based support provided by tutors and technical administration ensures a smooth operation of the e-learning course. The more complex a learning course the more support it needs. However, support and administration can be reduced if training courses include collaboration tools, where students can interact and exchange views, ideas and information. Newsgroups and mailing lists are a very economic way of using portals of this kind.

Types of evaluation

Evaluation research deals with several different types of evaluation. Each has its own suitability to assess the quality of projects and programs and make recommendations for improvement. In the following sections, various types of evaluation will be outlined. However, the focus will be on the suitability of evaluating learning arrangements in the field of web-based learning.

Formative and summative evaluations

Formative evaluations focus on the training course during its development. These types of evaluation aim at ensuring quality and provide useful suggestions for further improvement and refinement of the online course. Summative evaluations however examine the training course after it has been finalized. Here, the main focus is on data compilation. These data give useful hints regarding acceptance, effectiveness and the impact of online training courses.¹³

¹³ See Kromrey, 1998, p.100

Product-evaluations and process-evaluations

Product-evaluations consider one specific product when assessing (e.g. a learning program). In contrast, process-evaluations focus on procedures, handling and utilization of these products.

Internal evaluation and external evaluation

Those who have been actively involved in the development of the online-course do internal evaluation. They conduct the assessment and evaluation of its performance. External evaluation involves external assessors as a main resource to conduct the evaluation.

These types of evaluation provide a grid that should support the decision-making process, showing what type of evaluation is most suitable for which specific training course. Here, suitability depends very much on the type of training course, the state of its implementation, its composition, the assessor's perspective and most importantly the reasons for conducting the evaluation. Obviously a "one-size-fits-all" solution is not possible.

Evaluation methods and gathering of data

Evaluation criteria and empirical data are two central elements of every evaluation. Without evaluation criteria, the acquisition of data can quickly turn into a wild collection of data without correlations, interaction and structure. On the other hand one can say that without any empirical data, questions and presumptions remain without an answer.

In an evaluation, empirical data provide the foundation to validly answer theory-driven questions, to get clarity on assumptions and hypothesizes and to be able to

make recommendations. That is why when evaluating online training courses, methods of empirical research are necessary in order to help collect data without interfering with the actual learning process. Generally, internet-technology provides a number of possibilities to collect empirical data. This is especially relevant for written assessment methods (e.g. questionnaires, rating scales, etc.) that are applied frequently in evaluations.

Like all methods of empirical research, evaluation research can be divided into “reactive” and “non-reactive” procedures. When conducting “reactive” assessment procedures (such as interviews) the interviewee is aware of the assessment being conducted. Therefore the person assessed can react to the assessment process in an unpredictable way. The answers can mingle with the originally intended aim of the assessment in such way that it is difficult to make a distinction of all results obtained afterwards. “Non-reactive” assessment procedures (such as hidden observations) take place without the awareness of the subject of assessment. The reactive element of responding does not exist and therefore also no distinction of data is necessary after the assessment has taken place. As a consequence one can say: The assessment with reactive assessment procedures causes a wild mix of (interesting and not interesting) data.

Nevertheless, the utilization of non-reactive assessment procedures also has its challenges: After the assessment is finalized the subject of assessment should be informed on the objectives of the evaluation as well as the reasons for conducting the assessment in such a way. Generally, anonymity of all gathered data in reports should be guaranteed.

Therefore, when selecting evaluation methods it must be noted that obtaining empirical information depends on the decision whether to utilize reactive or non-reactive methods of assessment. A substitution is not possible.

Analysis of documents

Text and documents (e.g. curricular text, teaching text, documentation of communication amongst the participants including the mentor, etc.) on various levels play a crucial role within learning arrangements involving ICT. It is self-evident that all documents used in the learning context need to be carefully tested. When dealing with online training courses this issue becomes even more relevant and higher standards need to be set. This is because the mentor, who functions as a corrective element in the learning process, is not always directly available. It is advisable to make use of programmes designed for text analysis to analyse comprehensive text.

Interviews

Interviews are the most common method of data collection. Here, various variations of data assessment exist. The most popular distinction is made between oral interviews and written interviews. Oral interviews are mostly distinguished by their degree of complexity and structure.

- “Structured interviews” are based on a guideline that has been prepared beforehand. This guideline contains all relevant, necessary and already formulated questions to be asked during the interview as well as (if necessary) hints and tips with regard to the behaviour of the interviewer.

- “Semi-structured interviews” are based on clusters and groups of questions and topics to be dealt with during the interview. A specific sequence or wording of questions is not part of the guideline.
- “Freely structured interviews” are completely free of structure.

Written interviews or questionnaires can be divided based on the form of question, which is used in the interview.

- “Open questions” provide the interviewee with the possibility to formulate the answer in an individual manner.
- “Closed questions” suggest answering options to be chosen from.

Online-questionnaires should be the most common way of assessing and evaluating online learning processes.

Observation

Observations do not play a significant role when assessing online learning. Behaviour can be validly recorded through technology-based recording of behaviour via the usage of the learning platform.

Recording of behaviour

Behavioural expressions in the context of e-learning can be recorded via the so-called “log files” to be found through the server, where all html-documents of an online course are stored. These are access data that provide useful information for the evaluation process (e.g. acceptance of specific learning content, etc.). Besides that, “log files” inform about time, sequence and duration of utilization. However, one should not overestimate the role and function of these files and the information

provided. On the one hand they do not comprise all of the information necessary to conduct an evaluation.¹⁴ On the other hand they only provide clustered quantitative information that only might be of limited value for the overall assessment. After all, “log-files” only cover the access to HTML-documents. If other forms of communication like collaboration tools, bulletin boards are being used as part of the learning process it will not be captured by these files.

Testing

In evaluation research the term “testing” describes a standardized procedure to assess the occurrence of empirically defined performance characteristics. Usually assessment via “testing” is done in an ad hoc way and rather informally. Standardized forms of testing however are highly sophisticated and complex (e.g. intelligence tests). In standardized testing one distinguishes between “norm-oriented” and “criterion-oriented” methods.

- “Norm-oriented” methods assess an individual test result compared to a control group.
- “Criterion-oriented” methods are based on a predefined figure to assess individual test results.

Empirical research

When assessing online learning, methods of empirical social research need to be adopted and adjusted according to the specific need and demand created by the training course. This means that the whole range of research methods is of relevance for assessing online learning. Just as in any other area of empirical

¹⁴ If passwords have been given out, it will not be captured by “log files”.

research the decision on the most suitable method depends on the research question as well as the capacity of the respective research method.

Concepts of evaluation for web-based learning

So far we have dealt with evaluation of web-based learning from a methodical point of view. The following sections will apply the outlined methods with the aim of illustrating three types of evaluation concepts. Each one has aspects relating to data survey and data assessment. In other words, after dealing with evaluation of online-learning on an operative level, the following section will provide a “bird’s eye view” of the topic. The first two sections will illustrate aspects of data survey with regards to concepts of evaluation for web-based learning

Utilization of criteria indices

In the field of learning software, “criteria indices” are widespread. Generally, criteria indices can be described as checklists. Evaluation via criteria indices is based on a selection of various relevant and non-relevant criteria that have been pre-determined by experts. Due to their low-cost implications and easy application throughout the whole evaluation phase as well as transparency, criteria indices are very popular. Methods like these are relevant to obtain prompt results and to get a preliminary orientation for the overall implementation process. Furthermore, results gained through criteria indices are easy for others to understand.

However, they also have risks and challenges. These indices are often not based on sound theoretical ground. This shows in uncertainties when selecting criteria as well as the emphasis on each and every criterion. Furthermore, it has been empirically proven that assessors who utilize criteria indices when assessing

learning programmes obtain results that may deviate relatively strongly from each other.

Despite all criticism, one should not generally object to the method of criteria indices. They provide a grid for the area of web-based learning and a possibility for empirical research and evaluation. When expanding these indices through variables (e.g. drop-out rate), one can gain a proper instrument, with which online-seminars and web-based learning programmes can be validly evaluated. To extend this approach with research on students (e.g., individual self-assessment on learning progress) provides further possibilities of application that will be illustrated in the following section.

Determination of coherence

A step beyond the application of criteria indices are those concepts of evaluation, that not only deal with the existence or non-existence of characteristics, but also touch on the elaboration of coherence amongst characteristics. However, this is not free from difficulties. If coherence has been determined (e.g., between the amount of participating students and contributions in online discussions), or differences (e.g., difference between achieving a learning objective and learning groups), the finding is almost impossible to predict for other online-learning programmes if no other variables such as motivation of students are rigidly controlled.

Just as with criteria indices it might also make sense to analyze coherence relations via surveys involving training participants. In fact this is a requirement for many criteria and cannot be replaced by expert surveys.

Evaluations via criteria indices make it possible to allow statements on the theoretical impact of e-learning. A continuation of this approach leads to so-called

“linear structure equalisation models”, where a specific variable (e.g. estimated success in learning) will be determined from interrelations with other variables.

Chart 2: Concepts of evaluation

	Evaluation via criteria indices	Evaluation via analyzing relations
Perspective of the expert	<ul style="list-style-type: none"> • Learning success • Drop-out rate 	<ul style="list-style-type: none"> • Relation between learning success and acceptance of a specific learning/ teaching method
Perspective of the user	<ul style="list-style-type: none"> • Self-estimated learning success • Self-estimated degree of communication and interaction with other participants 	<ul style="list-style-type: none"> • Estimated relation between the degree of communication with other participants and own learning success

Aspects of assessing data

If the data assessment is based on criteria indices or interrelations, data alone only partially gives a statement on the quality of online-learning (including suggestions for improvement). The following section will provide an insight into various types of data that can be obtained when evaluating e-learning:

Data indicating learning success

A quantification of learning success only becomes a valid empirical statement through comparison (e.g. before/after-comparison) or through inclusion of analyzing relations (e.g. connection between learning success and participation in a specific learning module/ qualification module).

Decisions of participants/ experts

Judgments made by participants or experts are a fundamental empirical finding when evaluating online-learning. Before making changes based on such statements the data need to be examined further in order to ensure reliability and interrelations with other data.

Data relating to technical features

As long as these data are not utilized in relation with other relevant data of the evaluation (e.g. drop-out rate, learning success), these data can be considered as insignificant.

Data relating to the acceptance of the offered learning programme

Except for the case when responses in this regard are not accessible data, of this kind proves to be very difficult to interpret. Assuming that out of several alternative learning modules only one is accepted by a single learner or small group, it can still be of significance and relevance for a specific learning gproject.

Conclusion and discussion

Despite remarkable work being done in related fields such as educational software, evaluation research in ICT is still very much in its initial steps. However, one can assess online-learning on a practical level as long as an adequate system of classification comprising all relevant and necessary evaluation aspects has been developed. The system presented in the present paper is in conformity with these functional requirements as it provides a pragmatic, criteria-based evaluation that focuses on interrelations and thus serves the purpose of verifying or falsifying hypothesis-based evaluations.

References

- Arnold, Rolf; Schüssler, Ingeborg (1998): *Ermöglichungsdidaktik. Erwachsenenpädagogische Grundlagen und Erfahrungen*, Schneider Verlag, Baltimore.
- Heitmann, Werner (2004): The action-oriented learning approach for promoting occupational performance and employability. South African-German Development Co-operation, Skills Development Research Series, Book 3, Pretoria.
- Kromrey, Helmut (1998): *Empirische Sozialforschung*, 8th edition. Leske + Budrich, Opladen.

The Problem of Free Will in Program Evaluation

Michael Scriven

A group of hard-nosed scientists who have been studying the major commercial weight-loss programs recently reported their disappointment that the proprietors of these programs refuse to release data on attrition. The evaluators, though that's not the label they use, think it's obvious that this is a—or perhaps *the*—key ratio needed to appraise the programs, and one that the FDA should require them to release. On this issue (possibly for the first time in my life), I find myself taking sides with the vendor against the would-be consumer advocate, and I think the issue has extremely general applicability. My take is that the key issue is whether the program, *if followed*, will produce the claimed results; and that following the program is (largely but not entirely) a matter of strength of will. Failure to stay with the program—that is, attrition—is therefore (largely but not entirely) a failure on the part of the subject not the program, and the program should not be 'charged' with it.

First, here's why I think this is a very general problem that we need to deal with, in evaluation overall, not only in program evaluation. Think about the evaluation of: any chemical drug abuse program; twelve step programs like AA for alcohol and gambling abuse; distance or online education; continuing education of any kind—this clearly applies to all of them. Now it also applies in some important cases outside program evaluation, ones that you might not think of immediately. Here are two: (i) it applies to standard pharmaceutical drug evaluation because there is a serious problem referred to as the fidelity or adherence problem, about the extent

to which patients ex-hospital do in fact take the prescribed dosage on a regular basis. In these studies we surely want to say that the merit of the drug lies in what it does if it's used, not whether it's used. Case (ii): in teacher evaluation, although we want to say that the teacher has some obligation to inspire interest, to motivate, as well as to teach good content well, success is clearly limited, not only by natural capacity—as we all agree—but also by dogged disinterest. We don't want to blame teachers for failing to teach inherently capable students who are determinedly recalcitrant, i.e., for high failure ('attrition') rates where the cause is simply refusal to try.

Here's the schema I recommend for dealing with this kind of consideration. Think of a program (or drug regimen, or educational effort) as having three aspects that we need to consider in the evaluation: (A) Attractive power; (B) Supportive power; (C) Transformative power. For short: Appeal, Grip, and Impact. A is affected by presentation, marketing and perhaps allocation, and controlled by selection. The vendor or provider has the responsibility to use selection to weed out cases who are demonstrably unsuitable for the treatment; but, given the unreliability of such selection tests in the personnel area (pharmacogenomics is the subject devoted to this in the pharmaceutical area, where it's considerably more successful) and the importance of giving people a chance when they want to try, one can't be very critical of high-pass filtration for weight-loss, distance ed, and twelve-step programs. Of course, high front-end loading of payments may be excessive, if there's no money-back guarantee.

B is affected by support level including infrastructure (e.g., equipment, air conditioning, counseling), continuing costs (including opportunity costs and fees), and ease of use, for all of which the program is largely responsible; but of course B is also controlled by strength of will. If the support, costs, and ease of use are

disclosed in advance and are both reasonable and delivered as pictured and promised, willpower becomes the controlling variable. Which leaves C, the Impact issue, the real kick in the program: will it deliver as promised if we do our part, taking the pill, doing the homework, getting to the meetings? That's the key issue. While the good evaluator absolutely must check to see if the provider has indeed provided what was promised, and that what was provided was about as good as can be provided at the cost level in question, the rest is up to the subjects. Under these conditions, easily checked and often met, attrition is your failure, not the vendor's.

This is an important issue because it's important that evaluation not assume that these treatments are done to people, and are at fault if they don't work. The fact is that they are selected by people as something *they will undertake, not undergo*, and failure is often the fault of the people not the program. Even with drug treatments, the drugs have to be taken, and often taken for the rest of your life. They only work if you make them work. This is not surgery, which you *do* undergo, which *is* done to you; it's something where you choose to get some help in doing something to yourself. You have to take responsibility for doing your part, and the evaluator must not take that responsibility away and say that the program failed if it didn't get you through to the Promised Land, when it was you who failed. We have free will, but that doesn't mean *success* is a free lunch. Free will is the freedom to start a program: will power is what it takes to complete it.