MARTIN J. EPPLER* & FRANK MICKELER**

THE EVALUATION OF NEW MEDIA IN EDUCATION: KEY QUESTIONS OF AN E-LEARNING MEASUREMENT STRATEGY

The use of new media in education has moved from a minor field to an important application of new media. The measurement of the success and impact of e-learning applications is, however, still an unresolved research issue. In this article, we analyze the current status-quo and propose diagnostic key measurement questions as a strategic measurement planning framework.

Key Words: e-learning, new media, measurement, evaluation, training controlling, assessment, planning

*Institute for Media and Communications Management, University of St. Gallen, CH, martin.eppler@unisg.ch
**Institute for Media and Communications Management, University of St. Gallen, CH, frank.mickeler@unisg.ch

1. Introduction: The Measurement Gap in E-learning

As new media are rapidly becoming a central element in many learning and training contexts, the question arises whether the often considerable investments in e-learning applications have proven worthwhile and delivered the expected results or not. Consequently, *measuring* the quality and the impact of new media in education has become a crucial research and industry topic.

Two questions have dominated the measurement debate in this context so far. The first question that institutions which have invested in e-learning are asking is: are we measuring the right aspects (e.g., what should we measure to evaluate e-learning)? The second central question is: are we measuring these aspects right (e.g., how should we measure the benefits of e-learning)?

In this article, we show that in order to address these two questions of 'what to measure' and 'how to measure' systematically, a number of other questions must be asked. These questions go beyond the measurement objects and methods. Specifically, we will show that there are at least six diagnostic questions that any e-learning measurement strategy must answer in the planning stage. These questions do not only address the issues of *what* should be measured and *how* to best measure it, but also *why, by* and *for whom, when* and with what consequences the measurement should take place.[1] We hope that by introducing our framework, e-learning can be evaluated more consistently and systematically.

Before we present the framework in the main part of this article, we briefly review some of the central approaches towards training measurement in general and e-learning measurement or evaluation in particular.

2. The Status Quo: Kirkpatrick, Bloom etc. and the Question of What to Measure

Our framework is based on previous conceptual contributions that aim at distinguishing relevant areas of training assessment. They range from traditional training evaluation frameworks to e-learning assessment models. The most widely used framework in this domain is that of Kirkpatrick's

---

[1] These interrogative elements of our framework are taken and adapted from the work of Inglis et al. (1999). These authors state that the questions of what, when, and how to evaluate (and by whom) are not obvious and need to be addressed systematically.

four levels of training evaluation (see Kirkpatrick, 1994). Because of this, we will briefly review it below. Another important model of learning success is Bloom's taxonomy (Bloom, 1956) which we will also briefly discuss. Finally, we present in overview a typical e-learning assessment schema, that of e-learning quality by Ehlers (2002).

2.1 Kirkpatrick's Four Levels of Training Evaluation

Donald Kirkpatrick developed a model for the evaluation of training, which can be easily adapted for e-learning assessment. The model concentrates on the *output* of training and evaluates training benefits on four levels (see Kirkpatrick, 1994).

Level 1 ("Reactions" or "Satisfaction") focuses on the learners who participate in a training program and how they react to it. In this sense it is a measure of *customer satisfaction*. Positive reactions are important, because a negative learner perception towards the training experience makes it less probable that the newly acquired skills are transferred into the daily work, and hence generates a bad image among the workforce and the management. Although positive reactions may not ensure learning, negative reactions almost certainly reduce the possibility of insight or behavior changes.

Level 2 ("Learning") focuses on the *actual learning* that has taken place in a course or on the new or improved knowledge, skills or attitudes as the primary aim of the training event. This level can be measured through formal assessment and exams built in the e-learning system, and a later analysis e.g. looking at the retention rates some time after a course is completed.

On level 3 ("Behavior", "Application", "Performance"), the key question is: "are people *performing* better?" If a learner employs the new knowledge appropriately, the work behavior will meet certain desired criteria. According to Kirkpatrick, four conditions are necessary in order for a change in behavior to occur: The learners must have the desire to change, they must know what to do and how to do it, they must work in the right climate, and they must be rewarded for changing (see Kirkpatrick, 1994). The latter two conditions are independent of the training experience and are influenced by the organizational context. This shows that the success of an (e-) learning initiative also depends on other factors than the program or application itself, and that the best program is worthless when the organization is unable to accommodate the new knowledge.

The business results as the outcome of learning at level 4 ("Business Results") can be defined as the final results that occurred because the participants attended a training program (see Kirkpatrick, 1994). Therefore, a level 4 evaluation examines the contribution of e-learning to the *effectiveness of business*, showing whether the training results in more sales, fewer mistakes, higher productivity, or other similar measures.

Phillips (1996) separated the return on investment from level 4 and introduced this measure as a new level 5 (labeled "ROI"). Using level 4 evaluation data, the results are converted into monetary values and then compared with the cost of the training program (see Strother, 2002). An *ROI* analysis can be a lengthy and complex process, as the cost and benefits must be expressed monetarily, and all of the "soft" measures (e.g. better corporate climate, decreased personal contacts), have to be quantified.

It is to some extent critical to the concept that Kirkpatrick constructs a "chain of effect", which links each stage to the next level. Each level of evaluation depends on the previous one (see Setaro, 2001), and as one "moves from one level to the next, the process becomes more difficult and time-consuming, but it also provides more valuable information" (see Kirkpatrick, 1994). E-learning may in this sense offer the possibility of applying Kirkpatrick's schema more rigorously than ever before, because more measurement data can be made available and used for every level. Nevertheless, Kirkpatrick's framework is purely outcome focused and does not propose to measure any of the vital input factors, such as infrastructure, teacher qualification or content quality (see Abernathy, 2001). This is also true for the next system, Bloom's taxonomy, which classification of learning output or impact. It is briefly described below.

2.2 Bloom's Learning Domains and Taxonomy of Skills Levels

In the 1950's Benjamin Bloom proposed a hierarchy of learning behaviors categorized into three interrelated and partly overlapping domains covering *cognitive, affective* and *psychomotor* skills. The psychomotor domain refers to the use of mostly manual skills, i.e., coordination, and physical movement. The affective domain addresses a learner's emotions towards learning experiences. Attitudes, interest, attention, and values are demonstrated by affective behaviors. More important for e-learning is the first domain, as most electronic learning addresses the cognitive area. Within this area, there are two tiers which are composed of three levels each (see Taylor et al. 2002).

Tier 1 can be called "surface learning" and consists of the *knowledge* level, which measures the recall of information, the *comprehension* level, which requires a low level of understanding what has been taught, and the *application* level. Being able to apply what has been taught means being able to use general ideas, concepts, or methods in specific situations by transferring the knowledge from one context to another.

In tier 2 ("deep learning"), the taxonomy continues with the *analysis* level, which examines the learner's ability to break the learning content into its components, look at the relationships and interactions, and relate them to a common structure. On the *synthesis* level, a person is able to combine single elements into an integrated whole. The last level is that of *evaluation*, which means being able to make a judgement about the value of ideas, solutions, or methods. While Bloom's approach does not take into account the organizational impact of learning, he achieves a higher granularity than Kirkpatrick in terms of the actual learning quality as an outcome of training. In Bloom's view learning does not equal learning, as there are various levels of insight. This is an important aspect for e-learning testing applications, where various examination techniques can be applied to test the different levels of acquired knowledge, from mere recall to more profound synthesis or reflective evaluation. But again, the measurement of prerequisites and of the learning process itself is neglected. These aspects are addressed in the more recent framework by Ehlers (2002) that is presented in the next section.

2.3 Ehlers' E-learning Quality Levels

The three levels of e-learning assessment proposed by Ehlers (2002) are based on the premise that the notion of quality can be applied fruitfully to e-learning. Taking into account the *product view*, the *learner view*, and the *production view* of e-learning, quality in the context of new media in education means that an e-learning application achieves certain minimal standards in all three views, that it is free of errors, and that it is useful for the learners and provides an added value. For this to happen, one must, in Ehlers' view, look at three distinct levels.

First, the *prerequisites* for effective or high-quality e-learning (such as the computer infrastructure or the qualification of the tutors), second the *learning process* itself (including the learning culture in the company, the learning content, and the qualification goals), and third the *impact* or result of the learning in terms of increased problem solving ability. Ehlers

stresses the need for a global view on e-learning quality that not only fo-
cuses on the pedagogic or didactic aspects, but also on the economic and
societal ones (e.g., does e-learning foster lifelong learning and re-skilling).
While these aspects cannot be directly measured, they should be at least
considered when designing and evaluating an e-learning application. Ac-
cording to Ehlers, this global view of e-learning measurement or evalua-
tion should also lead to a measurement strategy that aims at evaluating an
e-learning program already in its *planning* and design phase, as well as in
its use phase and in its ultimate impact. Unlike the previously discussed
two frameworks, Ehlers does also address the question of *when* to measure
and *how* to measure, as he proposes various methods to measure the qual-
ity of an e-learning program in its various stages (Ehlers, 2002, p. 15).

We have incorporated the three levels of prerequisites/infrastructure,
process (incl. content) and impact or result into our framework and use it
to answer the question of *what* to measure. We have also added Ehlers'
idea of measuring *before* and *during* the actual deployment of an e-learn-
ing program (e.g., needs, expectations, acceptance etc.) in our *when* to
measure section (see the next chapter). While Ehlers' framework ad-
dresses many of the operational issues involved in e-learning measure-
ment, he does not address the questions *for whom* the measurement
should take place (he only focuses on the learners themselves) and *what
should be done* with the results of the measurement process (he mentions
quality improvement as the main measurement goal). Another question
that he does not address explicitly is *who* should be responsible for which
measurement activities. This and other questions will be addressed in our
framework.

2.4. Conclusion of the Framework Discussion

As we have seen, the models discussed above do not perfectly fit the cur-
rent organizational context of e-learning initiatives because they focus
only on selected measurement issues (e.g. what to measure or when to
measure). They either neglect the specificities of e-learning or the com-
plexity of measuring learning impact. They do, however, provide insights
into relevant measurement areas and feasible indicators. In our frame-
work, we will thus use them mostly in the 'what to measure' and 'when to
measure' section and we will use a variety of other sources for additional
measurement elements. The framework of e-learning measurement is
presented in the next section.

3. The Key Questions Measurement Framework

Having briefly reviewed some typical views on training and e-learning measurement, we can now introduce our framework of key questions for e-learning measurement. The framework is based on the premise that a consistent e-learning measurement strategy needs to address all interrogatives before starting to invest in e-learning measurement. Besides the three measurement frameworks discussed in the previous section, we use a number of other approaches. They are referenced in the introductory paragraph to every key question. First however, we define the scope of the framework and present its main elements in overview.

3.1 Scope of the Evaluation Framework

Before an e-learning system is introduced in an institution, much development work is necessary and many decisions need to be taken, as even the standard products and solutions can't be implemented ignoring the company context for which they are intended. The following types of evaluation serve for e-learning quality assurance from planning to the roll-out (see Inglis et al., 1999):
- evaluation of needs (corporate skill gaps, education needs, intended users, genuine demand, etc.),
- market analysis (available software products, service providers, best practices, etc.),
- design evaluation (test of ideas and assumptions during the concept phases),
- formative evaluation (during the project in order to monitor progress and modify developments as it proceeds).

The focus of our work in this paper, however, lies on the *summative assessment* of an e-learning system after the implementation (although evaluation activities need to be conducted already in the planning phase). Summative or program evaluation is suitable for looking at the success of a project, and serves as an opportunity to improve the quality of the e-learning environment – thus better meeting the needs of learners (see May, 2000), and also the requirements of top management. Our framework is thus less suited to evaluate e-learning products or e-learning suppliers. It is rather a management framework that can be applied by corporate e-learning managers to be used in their organizational setting.

3.2. The Framework in Overview

Figure 1 depicts the framework for a comprehensive e-learning measurement strategy in overview. It shows that there are six main questions which must be answered before investments in e-learning measurement infrastructures and processes are made. These questions relate to the main parameters of e-learning measurement and are as follows:

The first question relates to the main objectives of the measurement, that is to say *why* a company decides to measure its e-learning program. Here, we distinguish between strategic reasons and operational reasons. Strategic reasons are e.g. investment decisions related to e-learning. Operational reasons are mostly decisions regarding the improvement of an e-learning system. The second question which needs to be addressed (and which is closely related to question one) is *for whom* the measurement is to be made. Again, we distinguish between strategic stakeholders, such as top management, and operational stakeholders, such as e-learning managers or the learners themselves. The third question relates to the measurement object or *what* is actually measured. Here, we distinguish between input and output measures. Input measures relate to the e-learning prerequisites, context, content, process and staff etc. Output measures focus on the impact or effect of e-learning (such as learner satisfaction or changes in their behavior). These first three questions define the *context* of an e-learning measurement strategy. The second set of questions is more applied and relates to the *implementation* of e-learning measurement based on the previously defined context. The fourth question is thus *how* to measure, e.g., which tools and methods can be used to quantify the input or output measures (such as surveys, logfiles, or usability labs). These measurement tools must be administered by qualified people. The fifth question of *who* should measure addresses this point. The final question determines at what time the tools are used by the responsible staff, that is to say *when* to measure. Here, the options are a prior measurement before the e-learning investment, during the actual e-learning use, or afterwards (immediately after an e-learning training or at a later stage). Alternatively, one may decide to measure e-learning on a continuous basis. These questions and their relationships are depicted below.
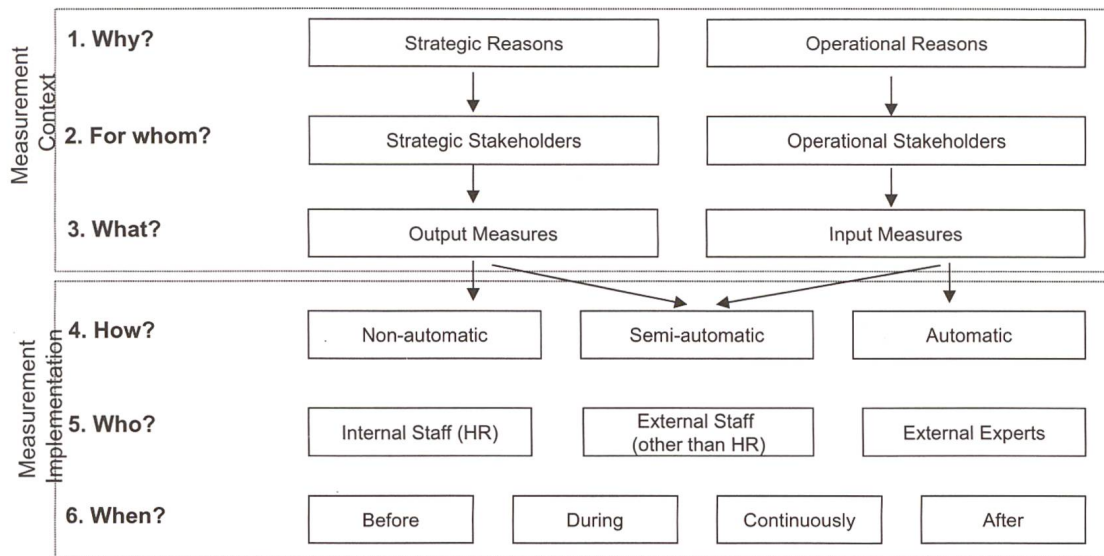
Measurement Context			
1. Why?	Strategic Reasons		Operational Reasons
2. For whom?	Strategic Stakeholders		Operational Stakeholders
3. What?	Output Measures		Input Measures

Measurement Implementation				
4. How?	Non-automatic	Semi-automatic		Automatic
5. Who?	Internal Staff (HR)	External Staff (other than HR)		External Experts
6. When?	Before	During	Continuously	After

Figure 1: The e-learning Measurement Framework

Next to these six main questions are other vital aspects of measurement which we haven not included in the core elements of the framework. These additional questions (see section 3.4.) regard the effects of the measurement process (i.e., who should take which kinds of actions based on the measurement results) and the insights gathered about the measurement process (e.g., how the measurement can be improved in future evaluations). The six main questions and these additional questions will be addressed in detail in the next sections.

3.3 The Measurement Questions in Detail

3.3.1 Why should we measure ? (Measurement reasons or objectives)

Before an e-learning program evaluation can take place, one should take a close look at the target audiences of the evaluation results and how they will use the measurement information. Depending on the target group, different aspects and quality issues become important. A strategic reason to evaluate an e-learning effort can be an investment decision between competing follow-up e-learning projects, and therefore hard facts and numbers are generally preferable. From an operational point of view, the reason of evaluation is more directed at the ongoing adaptation and optimization of the learning system. In the table below, we list the main strategic and operational reasons for e-learning measurement that can be envisioned and that we have encountered in the relevant literature

(Phillips, 1996a, Wentling et al., 2000, Mann / Robertson, 1996, Shepherd, 1999a, Todesco, 1997).

Measurement Questions	Answers for e-learning
1. Why should we measure? (Measurement reasons or motives)	**Strategic Reasons** *Functional level (human resources department)* <u>reactive</u> • to justify the cost and show all benefits (hard and soft) of an e-learning application • as a reaction to reengineering and downsizing efforts, to maintain training funds (in times of corporate belt-tightening) • to demonstrate accountability (as training has become more visible in organizations) • as a means to compare e-learning to traditional training programs by using the same effectiveness measures <u>proactive</u> • to get support and facilitate planning for future projects • to show the improvements from (radical) new processes introduced • to determine the success in accomplishing a program and show its potential for the future • as a marketing tool (gather data to assist in marketing future programs) • to enhance training status *Corporate level* <u>to keep track / monitor:</u> • to identify the strengths and weaknesses in the HRD process (evaluation of and not by the HR department) • to calculate the return or measure the impact of an e-learning investment • to compare the cost to the benefits of an HRD program to demonstrate the increase in a company's return on training dollars • as a measure of the growth in intellectual capital as a non-financial (intangible) asset • as a means for public relations <u>to prepare decisions:</u> • to establish a database that can assist management in making training decisions • to decide who should participate in future programs (to decide who is likely to benefit most)

	• to determine if an e-learning program was the appropriate solution for the specific business need, to validate training as a business tool (compared against other methods for improving performance and profitability)
	Operational Issues
	To adapt the learning system
	• to improve or adapt e-learning applications in various dimensions (e.g. to test the clarity and validity of tests, the quality of tutoring etc.)
	• to help in selecting training methods (make rational decisions about the methods, e.g. classroom, self-study, on-job, to employ)
	To assess the learning system
	• to identify which participants were the most successful with the program
	• to gather user feedback on advantages and disadvantages of the system[2]
	• to measure the popularity and acceptance of the system

3.3.2 For whom should we measure? (Target groups)

Closely related to the question of why to measure is the "for whom"-aspect. Different people in different roles are interested in diverse aspects, quality levels or measures of outcome, depending on what they intend to use the information for. Again, it makes sense to differentiate between strategic and operational stakeholders. The former are mostly the investors and sponsors of e-learning, e.g., upper management and other key stakeholders. The latter comprise all individuals involved in the operation of the system, from the learners, the IT-support staff to the HR employees engaged in the administration of the curricula. The following table is also based on May (2000) and Baumgartner (1999).

---

[2] May writes the following on this topic: "Effective assessment procedures need to employ methods for feedback which alert the learner to areas in which they have a deficiency. By integrating opportunities for assessment in the total learning process, the information returned to the learner can lead to remediation during the learning experience" (May, 2000).

| 2. For whom should we measure (Target groups) | Strategic Stakeholders

Internal target groups:
• top management / board of directors
• middle management
• internal sponsors
• management of the HR department
• controlling / accounting department
• all employees
• workers' council

External target groups
• (potential) clients
• potential new employees
• investors and analysts
• rating or certification companies
• trade union representatives
• government or regulatory authorities
• professional associations and organizations |
| | Operational Stakeholders

Clients of the system
• users of the e-learning system
• superiors of learners (responsible for their employees)

Suppliers / operations staff
• system administrators
• tutors (teachers, subject matter experts)
• IT-staff (technical service)
• authors of content, editors
• multimedia producers involved in the maintenance process
• other HR employees |

3.3.3 What should we measure? Which elements should we measure? (Measurement scope, evaluation objects)

On the input side of e-learning, we look at the prerequisites for the e-learning program to be evaluated and on the process itself (Ehlers, 2002) by proposing five objects to be measured: the e-learning infrastructure (the IT system), the e-learning process, its content, the organizational context, and the factors related to the learners themselves (such as motivation). As far as output measures are concerned, we use Kirkpatrick's

levels, adapted to the specific needs of electronic learning. This key question has been discussed widely in the literature, which therefore offers a wealth of material on what elements to evaluate (see Bersin, 2002, Rosenberg, 2001, Hall & LeCavalier, 2000, Shepherd, 1999, Setaro, 2001, Taylor et al., 2002).

3. What should we measure? Which elements should we measure? (Measurement scope, evaluation objects)	Input
	Infrastructure
	- web performance, downtime of the system - web usability (ease of use, structuring, orientation, navigation) - interface design (intuitive usage, consistent, consolidated, clean, clear, appealing and understandable design) - features and tools (built in to support the learning process) - website hits - number of registered users
	Process
	- quality and amount of feedback provided to the learner (by the system and by human administrators or tutors) - subject matter expert support levels - fitness for use of curriculum delivery methods (match of the delivery method to the curriculum design) - mode of instruction (match to the learning styles and preferences)
	Content
	- information quality, perceived usefulness of information - use of media (reasonable use of text, pictures, audio, video, simulations) - learnability of content - additional course materials (e.g. print-outs)
	Organizational context
	- training budget per employee - compatibility of learning and regular work - image of HR department and training staff inside the company - commitment of the top management towards employee qualification and knowledge development

The learner

- numbers (of learners)
- geographic distribution
- staff / employee qualification levels
- individual competency profiles
- readiness for participation and interaction (2 perspectives: learner-to-learner, learner-to-instructor)
- user behavior (motivation and persistence)
- familiarity with IT (regular PC usage, web usage)
- learners' perception of training as opportunity or threat

Output

Level 1 – Reactions
- individual satisfaction (desire to participate in future e-learning courses; helps to gain organizational support / immediate feedback)
- climate improvement

Level 2 – Learning (see Bloom's Domains and Taxonomy)
- cognitive domain
  (e.g. exercise completions, certifications received)
- affective domain (e.g. sensitivity to other people)
- psychomotor domain (physical skills, machine handling)

Level 3 – Behavior
- organizational linkage (= extent to which learners mix)
- communication behavior
- conflict resolution abilities
- efficiencies with business tools
- decreases in occupational accidents

Level 4 – (Business) Results
- customer satisfaction (repeat sales, customer complaints)
- productivity / efficiency, workload per employee
- implementation time of tools
- period of vocational adjustment of new employees
- performance to schedule
- income received
- quality of outcomes
- corporate image
- employee morale
- absenteeism
- turnover
- lost-time injuries
- workers' compensation insurance claims

Level 5 – Return on investment (ROI)
- return on e-learning investment (benefits vs. cost)

3.3.4 How should we measure? (Measurement Methods)

Ever more measurement and evaluation can be done automatically, as the functionalities of e-learning (learning management) systems are being increased. However, in the end there is always a human effort necessary to interpret the results and take follow-up actions. Some critical issues associated with the measurement methodology that need to be taken into account are the following ones:

• use and evaluate a nonequivalent control group not receiving the e-learning intervention in order to identify any differences in learning outcomes.

• make sure that your instruments meet your organization's needs with respect to providing valid and reliable information – have the instruments been validated in an acceptable manner?

• closely related is the problem of attribution of impacts to training in an environment where many influences are at play, especially on the higher output levels.

Next to these issues, one has to make sure that the complexity of the measurement tools can be handled by the organization and its staff. In the table below, we provide an overview on the tools that can be used in e-learning measurement (see also Hall & LeCavalier, 2000, Saba, 2000, Todesco, 1997).

Measurement Questions	Answers for e-learning
4. How should we measure? (Measurement methods)	**Automatic** • on-screen tests / achievement tests • problem-solving exercises • use of the mouse or keyboard (for the psychomotor domain which is often not well suited to online measurement) • simulations results • automatic statistical analysis (traffic, logfiles) • navigation tracking • assessment of the usage of electronic equipment (if subject of the learning intervention)
	Semi-automatic • web / data mining • user surveys and analysis • surveys of end users

	• online multiple choice questionnaires (questionnaires using web page forms) • email analysis
	Non-automatic (expert analysis) • questionnaire analysis • expert reports • steering committee feedback • one-to-one interviews • focus groups with separate user categories • analysis of diaries kept by users to record the context and outcomes from system use ("learning logs") • analysis of learner self-reporting (learners keep count of what they do) • essays submitted by learners • observation, observer ratings (observers = usually managers and supervisors, trained in the evaluation system) • conversation and discourse analysis of meeting minutes, discussion forum entries, etc. • assessment using one-to-one chat or group chat • controlling / accounting methods (at least for levels 4 and 5 (ROI))

3.3.5 Who should measure? (Measurement Responsibilities)?

The question of who should measure not only relates to the responsibilities of e-learning evaluation, but also to the cost, e.g., whose budget will cover the often time-consuming evaluation procedure. Here, we can differentiate between internal and external evaluation staff. Depending on internal qualification time resources, one has to decide whether the e-learning measurement is performed with internal resources or with the help of specialized external consultants. As incorporating an assessment and evaluation system means an additional expense, collaboration and partnerships are most likely the best means to ensure an effective integration of assessment into an e-learning course (May, 2000). Many of the e-learning architectures available on the market today offer already some evaluation features. The table below outlines the range of possibilities.

5. Who should measure? (Measurement responsibilities)	**Internal Staff (HR / training)** • administrators • trainers • tutors • e-learning managers • the system itself (evaluation functionalities built into Learning Management software)
	External staff inside the company • employees of the controlling department • IT staff • general managers • learners (learner self-assessment)
	External people • universities • certification institutions • outside experts / consultants

3.3.6 When should we measure (Measurement Timing)?

There are many opportunities, when evaluative activities of a corporate e-learning effort are useful. Measurement can take place before, during or after an e-learning program, or it can be used continually.

To enable comparisons and calculate ROI or other benefits, it is necessary to evaluate both before and after the training. However, to detect mistakes or quality problems, a continuous monitoring may be necessary. The table below, which covers again aspects we encountered in the literature (May, 2000, Shepherd, 1999a, Worthen, 2001), shows these different possibilities of measurement timing.

6. When should we measure? (Measurement timing)	**Before** *Learner pre-assessment* • assess initial skill levels and -gaps • notify the instructor of the applicability of the material, and the need for additional remediation • measurement and evaluation to offering mass customization in learning, and to meeting the learning needs of the customer *Training investment analysis* • assessment of training results before training: forecast of monetary benefits that are likely to be gained from training

	During • formative learning assessment (deployed during the learning process) *After* • immediately after the training (especially level 1 and 2 evaluations) • later after the training (measure of "deep learning", e.g. knowledge retention, permanent behavior changes) *Continuously* • evaluation as an integral part of the e-learning efforts • especially level 3- and 4-data should be collected over a longer period of time

3.4 Other measurement questions

As mentioned before, there are other relevant (meta-) questions that can be asked about the measurement of e-learning. One crucial question that should be considered at the very beginning of an e-learning measurement process is what the results will be used for or which kind of changes can be expected in light of the gathered measurement evidence: Are these measurements limited to the e-learning system or can they affect the entire e-learning initiative? Are the follow-up actions only short-term measures or also longer term actions? Another key question in this context relates to the responsibilities for these measures: Who should take actions based on the measurement results? Is it just the e-learning staff or can the results also affect the way that learners use the program and management provides funding? These questions are best addressed in the planning stage of the measurement.

A final important question that should be considered throughout the measurement process relates to insights about the evaluation of new media in education: One should keep in mind what can be learned about the measurement process from the steps that were examined so far. This question ensures that there is reflection about the measurement process and how it can be improved in the future with regard to costs, quality and timing.

These questions are again presented in a checklist format below.

8. Who should take actions? (Follow-up Responsibilities)	• top management • e-learning managers • content authors • IT staff • content providers • learners
9. What can be learned about the measurement process for future measurement activities?	• reflecting on how to improve the efficiency of the measurement process • reflecting on the cost/benefit ratio of certain measurement areas (what hasn't been worthwhile measuring) • reflecting on the various roles of the involved stakeholders (e.g., how conflicts among them can be minimized) • reflecting on the interpretation of the measurement results, e.g., interpretation of indicators and their change over time

This concludes our survey of critical questions that need to be answered by an e-learning measurement strategy aimed at evaluating existing new media in institutional education. In the next section, we summarize our findings.

4. Conclusion

Up recently two questions have dominated the discussion on e-learning measurement, namely what to measure and how to measure. In this article, we have shown that in order to answer these questions, e-learning managers must first answer a number of related questions. These questions include the measurement motives, the target groups of the measurement information (e.g., who can use the measurement results and how), the adequate measurement timing, the measurement responsibilities and the measurement consequences. A crucial final question in our catalog of e-learning measurement questions relates to the learning about the measurement process itself and how the lessons learned can be used in the future to increase the validity and the impact of e-learning measurement. Future research in the area of e-learning measurement should explore the relationships between the questions that we have isolated in this article. This research could be conducted based on real-life cases of e-learning measurement initiatives and reveal how management choices and objectives affect the other parameters of the measurement process.

References

ABERNATHY, D.J. (2001) Thinking Outside the Evaluation Box, in: Training & Development, online: www.astd.org/cms/templates/index. html?template_id=1&articleid=21260 (25.9.02).

BAUMGARTNER, P. (1999) Evaluation mediengestützten Lernens, in: Kindt, M. (Ed.): Projektevaluation in der Lehre - Multimedia an Hochschulen zeigt Profil(e,, Münster: Waxmann, pp. 61-97.

BERSIN, J. (2002) Measure the Metrics, online: http://www.elearning-mag.com/elearning/article/articleDetail.jsp?id=21256 (15.10.02).

BLOOM, B.S. (Ed.) (1956) Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York ; Toronto: Longmans, Green.

EHLERS, U. (2002) Qualität beim E-learning: Der Lernende als Grundkategorie bei der Qualitätssicherung, in: Medienpädagogik, No. 02-1, online: http://www.medienpaed.com/02-1/ehlers1.pdf (14.10.02).

HALL, B., LeCAVALIER, J. (2000) The Case for Level 3, online: http://www.learningcircuits.org/nov2000/hall.html (04.10.2002).

INGLIS, H., LING, P., JOOSTEN, V. (1999) Delivering Digitally: Managing the Transition to the Knowledge Media. London: Kogan Page.

KIRKPATRICK, D. (1994) Evaluating Training Programs. The Four Levels. San Francisco: Berrett-Koehler.

MANN, S., ROBERTSON, I. (1996) What should training evaluations evaluate?, in: Journal of European Industrial Training, Vol. 20, No. 9, pp. 14-20.

MAY, K. (2000) e-learning; Improving Quality Through Assessment and Evaluation, online: http//www.eduport.com/community/kiosk/ 20004/ eval.htm (05.10.02).

PHILLIPS, J. (1996) ROI: The Search for Best Practices, in: Training & Development, Vol. 50, No. 2.

PHILLIPS, J. (1996a) Accountability in Human Resource Management. Houston: Gulf.

ROSENBERG, M. (2001) E-learning. Strategies for delivering knowledge in the digital age. New York et al.: MsGraw-Hill.

SABA, F. (2000) International Review of Research in Open and Distance Learning, in: IRRODL: Research in Distance Education, Vol. 1, No. 1.

SETARO, J. (2001) Many Happy Returns: Calculating E-learning ROI, online: http://www.learningcircuits.org/2001/jun2001/elearn.html (08.10.02).

SHEPHERD, C. (1999) Assessing the ROI of training, online: http://www.fastrak-consulting.co.uk/tactix/Features/tngroi/tngroi.htm (08.10.2002).

SHEPHERD, C. (1999a) Evaluating online learning, online: http://www.fastrak-consulting.co.uk/tactix/Features/evaluate/evaluate.htm (08.10.2002).

STROTHER, J. (2002) An Assessment of the Effectiveness of e-learning in Corporate Training Programs, in: International Review of Research in Open and Distance Learning, Vol. 3, No. 1, pp. 1-17.

TAYLOR, D., GOLES, T., CHIN, W. (2002) Measuring Student Learning, in: e-Service Journal, Vol. 1, No. 2, pp. 41-51.

TODESCO, A. (1997) From Training Evaluation to Outcome Assessment: What Trends and Best Practices Tell Us, online: http://learnet.gc.ca/eng/rescentr/fulltx/outpap.htm (10.10.02).

WENTLING, T., WAIGHT, C., et al. (2000): e-learning - A Review of Literature, Research Report, Urbana-Champaign.

WORTHEN, B. (2001) Measuring the ROI of Training, online: http://www.cio.com/archive/021501/roi.html (10.10.02).

Appendix – Selected Online Evaluation Frameworks and Guidelines

- American Society for Training and Development (various evaluation tools and links): http://www.astd.org/virtual_community/comm_evaluation/tools_links.html (15.10.02).
- Small Group Instruction Diagnosis: SGID [...] is a method of evaluation that uses facilitated small group discussion among students to provide feedback to an instructor for the purpose of improving teaching, which has been adapted for e-learning (Wentling / Waight et al., 2000): http://www.miracosta.cc.ca.us/home/gfloren/sgid.htm (15.10.02).
- American Evaluation Association: http://www.eval.org/ (15.10.02).
- The Learning Resource Network: Evaluating Pilot Courses: Does Your Course Design Work?: http://learnet.gc.ca/eng/lrncentr/online/pilotedemo/start.htm (15.10.02).