

Uranchimeg Tudevtagva

Structure Oriented Evaluation
Model for E-Learning

Wissenschaftliche Schriftenreihe

EINGEBETTETE, SELBSTORGANISIERENDE SYSTEME

Band 14

Prof. Dr. Wolfram Hardt (Hrsg.)

Uranchimeg Tudevdagva

**Structure Oriented Evaluation
Model for E-Learning**



**TECHNISCHE UNIVERSITÄT
CHEMNITZ**

Universitätsverlag Chemnitz
2014

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Chemnitz, Techn. Univ., Habil.-Schr., 2014

Technische Universität Chemnitz/Universitätsbibliothek

Universitätsverlag Chemnitz

09107 Chemnitz

<http://www.bibliothek.tu-chemnitz.de/UniVerlag/>

Herstellung und Auslieferung

Verlagshaus Monsenstein und Vannerdat OHG

Am Hawerkamp 31

48155 Münster

<http://www.mv-verlag.de>

ISSN 2196-3932 (Druck)

ISSN 2196-4815 (online)

ISBN 978-3-944640-20-4

<http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-146901>

Preface

The present volume of scientific series *Eingebettete, Selbstorganisierende Systeme* is dedicated to evaluation of e-learning.

The increasing complexity of economic, technological and social processes requires adapted management and control instruments. For future knowledge society this includes development of intelligent technologies, as well as adjusted methods for knowledge transfer. Thereby e-learning becomes a key technology of any education system.

The recent dynamic development on whole e-learning sector urgently needs an adequate quality management. For that corresponding evaluation models are required. This is the mission which Dr. Tudevdagva picks up in this work.

The focus is on the development of a logical structure oriented evaluation model for e-learning based on the general measure theory. Result is the structure oriented evaluation model (SURE model). This model extends evaluation methodology by including the logical structure of e-learning processes.

By consequent application of theoretical well defined measurement rules a consistent evaluation of embedded partial processes becomes possible. That is an additional result of SURE model which meets the varying expectations of involved groups in e-learning.

The SURE model of Dr. Tudevdagva is beside of e-learning also applicable for evaluation of other structure oriented processes like administration or organisation processes, for instance.

I am pleased that I could win Dr. Tudevdagva for publication of her SURE model in this scientific series. I am convinced that the pioneering work of Dr. Tudevdagva has opened a new branch in evaluation theory.

2.7A

Prof. Dr. Wolfram Hardt
Head of Professorship for Computer Engineering
Vice Dean of Faculty of Computer Science
Technische Universität Chemnitz

Reviewers

Prof. Dr. W. Hardt, Technische Universität Chemnitz, Germany

Prof. Dr. T. Köhler, Technische Universität Dresden, Germany

Prof. Dr. E. B. Tsoy, Novosibirsk State Technical University, Russia



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

Structure Oriented Evaluation Model for E-Learning

Habilitationsschrift

zur Erlangung des akademischen Grades

Doctor rerum naturalium habilitatis
(Dr. rer. nat. habil.)

vorgelegt

der Fakultät für Informatik
der Technischen Universität Chemnitz

von PhD Uranchimeg Tudevdagva,
geboren am 11. September 1969 in Ulaanbaatar.

Chemnitz, den 27. Januar 2014



This work was performed during a Post-Doc Research at Technische Universität Chemnitz, Chemnitz, Germany.

The project was supported by Schlumberger Foundation, Faculty for the Future. The financial support of this project by the Schlumberger Foundation is gratefully acknowledged.

Acknowledgement

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me.

Above all, I would like to express my very special appreciation to Professor Wolfram Hardt for his great support, valuable and constructive suggestions. This thesis would not have been possible without his help and understanding during development of new ideas.

I am thankful to all members of Department of Computer Science of Technische Universität Chemnitz for their support during this research. I am very thankful to Dr. Ariane Heller for her help at integration of my family into new environment.

I wish to thank Professor Karl-Heinz Eger for his valuable suggestions during the preparation of this research work. I would like to extend my thanks to Professor Thomas Köhler from the Dresden University of Technology and Professor Evgeniy Borisovich Tsoy from the Novosibirsk State Technical University for their constructive interest.

I would like to express my thank to host University, the Technische Universität Chemnitz, where I found an excellent research environment. I also would like to thank the Mongolian University of Science and Technology which made it possible to perform this Post-Doc study abroad.

Last but not least I would like to thank my husband Oyunbat Tumurbaatar and my lovely children Gantogoo and Oyunchimeg for their support, understanding and great patience at all times.

Contents

1	Introduction	3
1.1	Motivation and challenge of research	3
1.2	The structure oriented evaluation model	5
1.3	Related works and evaluation models	9
1.4	Summary	16
2	E-learning framework and SURE model	19
2.1	The e-learning framework	19
2.2	The need of evaluation in e-learning process	23
2.3	The structure oriented evaluation model	27
2.3.1	Basic idea of structure oriented evaluation	27
2.3.2	The eight steps of SURE model	33
2.4	The tool for SURE model	39
2.5	Summary	42
3	Theoretical foundation of the SURE model	45
3.1	Logical structure and evaluation model	46
3.2	Calculation rules for scores	53
3.3	Estimation of scores for goal structures	58
3.4	Calibration rules for empirical scores	60
3.5	Simulation examples	67
3.6	Summary	74
4	Pilot testing of SURE model in evaluation process	77
4.1	Application software for engineering computation	77
4.2	Virtual laboratory for maintenance of computers	87
4.3	Construction gas station and liquid gas station	93
4.4	Practice-oriented introduction to computer graphics	98
4.5	Summary	107
5	The research results	109
	List of Figures	113
	List of Tables	114
	References	115

I Introduction

I.1 Motivation and challenge of research

In an era of new technological developments, e-learning has become a central issue in future developments of education systems. E-learning is widely being used in conventional, continuing and adult education and corporate training because of its flexibility, richness, resource-sharing, and cost-effectiveness. United Nations Educational, Scientific and Cultural Organization (UNESCO) statistics show that over 455 million people around the world had received education and training through the Internet in 2008 [58]. By survey of Babson Survey Research Group over 70% of universities in USA were providing e-learning courses, and more than 6.1 million university students were taking at least one e-learning course during the fall 2010 term, which accounted for over 31% of the total number of university students in USA [4].

The need of e-learning in education and training is confirmed also by the E-learning investment report : *Corporate training is a \$200 billion industry, of which \$26 million is represented by e-learning. In Europe, 51% of companies delivered one training session via e-learning to more than 50% of their employees. Corporations in Spain and the U. K. lead the way with 56% and 53% of learners, respectively, using e-learning. Part of the reason for global growth of e-learning is the increasing of wireless connectivity. For example, South Korea, which has one of the highest-rated education systems, aims to have wireless networks in all schools by 2015, when all curriculum materials will be available in digital form.* [17].

In this context, the call for adapted educational standards, corresponding quality evaluation and process management of e-learning tools is becoming even louder. Educational evaluation has a long historical tradition, but evaluation for e-learning is still being developed as a new branch of educational evaluation. The rapid development of computer and mobile technology has become the main motivation to change the teaching and learning methodology in e-learning. Therefore, relating with teaching and learning methodologies evaluation models and methods are required to be updated.

Many researchers and scientists developed different approaches and mo-

dels for educational evaluation. R. Tyler [57], [56], D. Kirkpatrick [27] - [29], M. Scriven [44] - [46], M. Alkin [1] - [3], J. Messick [33], [34], D. Stufflebeam [48], [49] and J. Phillips [38] - [41] are pioneers in educational evaluation theory. Other researchers such as M. Patton [35], [36], U. D. Ehlers [13] - [15], B. Khan [24] - [26], F. Colace [11], [12], P. Lam [31], V. Ruhe [43], E. Taylor-Powell and E. Henert [51] extended and continued this research.

However, no approach for evaluation of e-learning could reach a general acceptance until now. For a corresponding overview we refer to the report of Swedish National Agency for Higher Education [42]. This report contains an excellent survey on the European view on e-learning and quality assessment. Concise further overviews on quality research in e-learning are given by D. Ehlers [13], V. Ruhe and D. Zumbo [43]. For some additional or special aspects we refer to B. Khan [24], D. Kirkpatrick [27], D. Stufflebeam [48], P. Zaharias and A. Poulymenakou [60], E. Taylor-Powell and E. Henert [51].

Educators can use different evaluation models or methods. The main condition is that the evaluation process should be transparent and traceable for involved groups of the e-learning process.

The quantitative models for evaluation of e-learning usually used are additive evaluation models. That means, depending on the considered evaluation goals, which are measured based on a defined scale, a linear function containing corresponding weight factors is used like, for instance,

$$Q = \sum_{i=1}^r \alpha_i x_i.$$

Here denote α_i , $\alpha_i > 0$, given weight factors for the obtained measure values x_i , $i = 1, \dots, r$, for the considered evaluation goals. The advantage of additive evaluation models is that they are easy to use. A side effect of these models is, that the choice of proper weight factors is subjective. It is difficult to define which element should be more highlighted and should have highest weight factor.

According to our observations there is a gap here. The existing evaluation models usually have no clear data processing part. The planning of evaluation, to create survey or checklist, to report evaluation result, all these parts of evaluation process are extensively discussed in literature. But

it is not easy to find materials which explain in detail how the collected data should be processed.

By our view collected data have to be processed transparent and as objective as possible without any subjective influence. This was the starting point for development of our evaluation model which has clear calculation rules to process collected data.

The challenge of our research was to create and develop an evaluation model which fits to rapidly changing learning environment of e-learning. Aims of our research are:

- to develop an evaluation model that supports scientific evidence;
- to develop adapted calculation rules for the model;
- to test the model with simulated and authentic data.

Hence in this study, we develop an evaluation model which can be used by all groups which are involved in an e-learning project. Result of this research is a basic evaluation framework that can be collaboratively used by all groups of evaluation process.

1.2 The structure oriented evaluation model

The structure oriented evaluation model (SURE model) that is being discussed here is the result of a long term study of evaluation methods in e-learning, U. Tudevågva and W. Hard [52], [53]. Idea of this model is based on practical aspects of e-learning.

In early stages of e-learning, the stakeholders were the initiators of evaluation process. The reason was that the word 'evaluation' was understood as to measure and to rank out the different activities and projects. According to our understanding evaluation of e-learning has to support additional functions. Evaluation can be used as a tool or instrument for improvement of e-learning. Universities and other institutions who develop and offer e-courses need self evaluation. They can obtain dependable and trustworthy results by the SURE model. When the evaluation model is complete and transparent people will welcome the findings of such evaluation process. This is ensured by structure oriented evaluation model.

The SURE model consists of eight steps. All steps of structure oriented evaluation model have a specific meaning. Output of the previous step will become the input of next step [54].

These are the eight steps of SURE model:

- Definition of key goals
- Definition of sub goals
- Confirmation of evaluation goals
- Creation of checklist
- Acceptance of checklist
- Data collection
- Data processing
- The evaluation report

They include the following:

Step 1. Definition of key goals

During the first step of SURE model, evaluation team has to define the key goals of evaluation. The achievement of these key goals is essential for the achievement of overall objective of e-learning. If one of the key goals is failed, then the e-learning process will be evaluated as failed.

Step 2. Definition of sub goals

Here the key goals defined in the first step, can be described in a more detailed version. For that sub goals can be used. Sub goals can be interpreted as different ways to achieve a key goal. Hence the SURE model supports series and parallel goal structures.

Step 3. Confirmation of evaluation goals

This step is an important step which is often neglected or forgotten. The evaluation team has to confirm the final version of evaluation goals which has been defined during the previous two steps. When all members of the evaluation team have accepted the logical structure of evaluation goals, that

should be fixed by a protocol. This is helpful in avoiding future conflicts between evaluation team and involved groups of e-learning.

Step 4. Creation of checklist

The checklist is the most used type of data collection. This is a well developed data collection method and there are many commercial and open source software solutions to create online surveys or checklists. The checklist of SURE model has to be created based on confirmed goal structure. To create a checklist, one can use different online software. The weakness of existing software solutions is that they cannot take reference to logical structure of the evaluation goal. Moreover they don't include score calculation modules as they are required by the SURE model.

Formulation of questions is only one part of checklist. When the questions have been fixed evaluation team has to decide which type of a survey should be used for checklist. Depending on the target group of evaluation, design of the checklist can vary. The evaluation team has to focus its attention on this aspect carefully.

Step 5. Acceptance of checklist

When the checklist is ready, evaluation team has to confirm formally the checklist. The team also has to check whether the checklist is adapted according to the logical structure of evaluation goal, whether the questions are easy to read, understandable and no-dual meaning, whether the questions are clearly formulated, whether there are any grammatical and semantic errors in sentences. Only the accepted checklist can be used for data collection.

Step 6. Data collection

Data can be collected on different ways. If it is a paper based version, it can increase the subjective influence, because these data have to be tabulated by using human resources. If it is a web based online version, it can be more objective. There are several techniques for data collection: surveys and questionnaires, tests and assessments, interviews, focus groups, action plans, case studies, and performance records [41]. However, the SURE model should use an online survey.

Step 7. Data processing

One of the advantages of structure oriented evaluation model are the clear defined calculation rules for data processing step. This is the basic scheme of data processing: Let us consider an evaluation structure C which consists of r key goals $B_i, i = 1, \dots, r$, where each key goal is defined by s_i sub goals $A_{ij}, i = 1, \dots, r, j = 1, \dots, s_i$. We suppose that we have obtained n checklist data records via a checklist adapted to goal structure.

Let $[x'_{ij}, x''_{ij}], x'_{ij} \leq x''_{ij}$, be the evaluation interval of check list for evaluation how the aim of sub goal A_{ij} has been achieved.

The empirical score $Q^*(C)$ for evaluation of goal structure C is calculated by

$$Q^*(C) = \frac{1}{n} \sum_{k=1}^n \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)}) \right). \quad (1)$$

Here $q_{ij}^{*(k)}$ denotes for $k = 1, \dots, n$ the empirical score for sub goal A_{ij} according the k^{th} checklist result. This value is obtained by normalisation of checklist value $x_{ij}^{(k)}$, where $x_{ij}^{(k)}$ is the obtained score or answer of k^{th} student to checklist question how the sub goal A_{ij} has been achieved.

How this formula is obtained and how we have to interpret the empirical score $Q^*(C)$ is discussed in Section 3.

Step 8. The evaluation report

Evaluations are expected to contribute value to sustainable development. Understanding what works, what doesn't and what should be improved, promotes informed decision-making about programming choices, approaches and practices. Good evaluation reports serve this process by accurately distilling and clearly articulating what is learned from evaluations [22].

By SURE model the evaluation process consists of quantitative and qualitative steps. First five steps outcomes ensure qualitative part of the evaluation. Step 6 and 7 ensure quantitative aspects of evaluation. The evaluation report has to include both aspects of the evaluation process. The evaluation report has to make visible outcomes and findings of the evaluation process.

The type of report can be different depending on the audience to whom

that report is needed. By SURE model we can calculate several evaluation scores. Which evaluation score to be used in what, can be decided by the evaluation team. Graphics and charts can be produced automatically by the SURE model, and these graphical representations can be used in the evaluation report.

Evaluation report should be presented to stakeholders or audiences at appropriate times. Else all effort of evaluation team will be failed. As a rule all interested groups expect a quick report after data collection. Usually after collection of data until generating the final evaluation report there is a delay in fixing results and facts.

In frame of our research, we developed a beta version of a tool for SURE model. The architecture of this tool consists of five units.

- a)* The structure editor. This goal editor includes functions to create key and sub goal sets. The obtained logical structure is used in the fourth unit of architecture for data processing.
- b)* The checklist generator. Here user can enter checklist questions using the checklist editor. Result of this unit will be an online questionnaire for SURE model.
- c)* Online data collection tool interface. This unit serves the users during online survey. Administrator of the tool can control and check data collection process via this unit (Step 6 of SURE model).
- d)* Data processor. One of the advantages of SURE model are the adapted calculation rules for data processing (see Section 3). All collected data are tabulated and processed automatically (Step 7 of SURE model).
- e)* Graphic generator. This beta tool for SURE model includes simple functions to generate different types of charts and graphics (Step 8 of SURE model).

1.3 Related works and evaluation models

An evaluation of education is one of the main research directions in education systems. The reason is, education is the basic intellectual value in

each country. Educated people always were the essential 'engine' of development of a country in almost all sectors: politics, health, education, engineering, social sciences and technology. Not only schools and universities need educational evaluation, each organization need to improve knowledge and skills of its employees continuously. Stakeholders invest a large amount of money on educational programs and also on advanced training courses. As a consequence high quality, successfully running courses are expected as outcomes of these investments. Evaluation of education is one of the basic methods to measure this outcome.

According to the 'Historical development review in evaluation' by L. Hogan [19], there are seven significant time periods of historical development of program evaluation, from the 1700's until the 21st Century.

Period 1: The Age of Reform (1792-1900's)

Period 2: The Age of Efficiency and Testing (1900-1930)

Period 3: The Tylerian Age (1930-1945)

Period 4: The Age of Innocence (1946-1957)

Period 5: The Age of Development (1958-1972)

Period 6: The Age of Professionalization (1973-1983)

Period 7: The Age of Expansion and Integration (1983-Present)

Moreover L. Hogan created a list of Evaluation approaches [19]: Objectives-Oriented, Management-Oriented, Consumer-Oriented, Expertise-Oriented, Adversary-Oriented, Participant-Oriented, CIRO Evaluation, D. Kirkpatrick's Evaluation and P. Phillips' Evaluation Approach. Hence L. Hogan made a clear picture of historical development of program evaluation. His work is appreciated in evaluation literature available on evaluation research, because he included not only the historical development of evaluation research, but created a classification of evaluation approaches and made a prognosis for future development.

M. Scriven [47] gave a fundamental review on evaluation research. He mentioned three main revolutions in educational evaluation theory. They are as follows:

1. *The first revolution was to transform the status of evaluation from untouchable to respectable, about a century ago the value-free doctrine held*

the view that there could be no place for serious treatment of evaluation within sciences.

2. *The second revolution was to transform the status of evaluation from that of a respectable discipline to that of the alpha discipline.*
3. *The third revolution was to extend the status of evaluation from that of an alpha discipline to that of the a paradigm applied discipline.*

M. Alkin [3] created an 'evaluation theory tree' first time in evaluation literature. There are three main branches in this tree: use, methods and value. M. Alkin and C.Christie called this tree 'Family tree of evaluation theory'.

The European Commission published in 2006 a report on 'Evaluating e-learning' [5]. This report contains a bibliography on e-learning evaluation models. There the literature for e-learning evaluation is divided into ten categories:

- *Case studies of specific e-training programmes*
- *Comparisons with traditional learning*
- *Tools and instruments for evaluation of e-learning*
- *Return on Investment (ROI) reports*
- *Benchmarking models*
- *Product evaluation*
- *Performance evaluation*
- *Handbooks for the evaluation of e-learning*
- *Meta-studies*
- *Studies on the contribution of evaluation to meta data*

We now present a short survey on existing evaluation models.

1. *Objective model.* R. Tyler [56] created one of the fundamental works on curriculum by coining the term "evaluation" for aligning measurement and testing of educational objectives. Because his concept of evaluation consisted of gathering comprehensive evidence of learning, rather than just paper and pencil tests, R. Tyler might even be viewed as an early proponent of portfolio assessment. R. Tyler formalized his thoughts on viewing, analysing and interpreting the curriculum and instructional program of an educational institution in *Basic Principles of Curriculum and Instruction*.
2. *Four level of evaluation model.* D. Kirkpatrick [28] is known for the creation of a highly influential model for training program evaluation, consisting of four levels of training evaluation. D. Kirkpatrick's ideas were first published in 1959, in a series of articles in the *US Training and Development Journal* but later better known from the book he published in 1975 entitled, "Evaluating Training Programs". D. Kirkpatrick's four levels are designed as a sequence of ways to evaluate training programs.
3. *CIPP Model (Context, Input, Process, Product).* The CIPP Model for evaluation is a comprehensive framework for guiding formative and summative evaluations of programs, projects, personnel, products, institutions, and systems. The CIPP model of evaluation was developed by D. Stufflebeam [48] and colleagues in 1960's, out of their experiences of evaluating education projects for the Ohio Public Schools District. CIPP is an acronym for Context, Input, Process and Product. This evaluation model requires the evaluation of context, input, process and product in judging a programme's value. CIPP is a decision-focused approach to evaluation and emphasizes the systematic provision of information for programme management and operation.
4. *Provus's Discrepancy Evaluation Model.* The Discrepancy Evaluation Model (DEM) of M. Provus [23] which is reflected in the CIPP model, is an effective way to evaluate academic programs. M. Provus defined evaluation as the process of agreeing upon program standards, and using discrepancy between outcomes and standards to evaluate the program. As such, the DEM can be used as a formative assessment

to determine whether to revise or end a program. The model is ideal for finding problems by means of discrepancy. It can also be used to design a program from inception to conclusion, leading to summative evaluations.

5. *Eisner's Connoisseurship Evaluation Models.* E. Eisner's [16] connoisseurship model provides a strong framework and alternative to the experimental design evaluations. According to E. Eisner, connoisseurship is "the art of appreciation" and educational connoisseurs are those who have distinctive awareness of qualities in different settings. He thinks connoisseurs evaluate the issues by looking at their qualities and illuminate a situation so that it can be seen or appreciated by others as well. In the evaluation of educational technologies, the connoisseurship model has two major implications: holistic approach to the analysis and interpretation of data and multiple perspectives in the evaluative tasks.
6. *CIRO (Context, Input, Reaction, Outcome).* Another four-level approach has been developed by P. Warr [59] with team. Originally used in Europe, this framework has much broader scope than the traditional use of the term "evaluation" in the United States. Main idea of this model is: The training department, like all other departments, will be expected to play its part in the achievement of the organisation's objectives. If trainers can demonstrate factually that they are making a genuine contribution to the organisation's goals, this can lead to an increase in both the standing and influence of the training department within that organisation. The amount of support given by other members of the organisation will rest largely on the regard they have for the training department staff. So any activity which heightens that regard will ultimately benefit the training function.
7. *Goal-Free Evaluation Approach.* M. Scriven [45] defines goal based evaluation as any type of evaluation based on and knowledge of the program, person, or product. Focus is on goals and objectives. It measures how the project/program has done in achieving those goals and objectives. This evaluation approach can be used if the main purpose of evaluation is to measure planned outcomes. As the name indicates this evaluation approach begins without focusing on goals

and objectives of the program.

8. *Kaufman's Five Levels of Evaluation.* Further researchers, recognizing some shortcomings of D. Kirkpatrick's [28] four level approach, have attempted to modify and to add to this basic framework. R. Kaufman [30] has expanded D.Kirkpatrick definition of Level 1 and added a fifth level addressing societal issues. Then he gets five levels: Level 1- Input and Process, Level 2 and 3 - Micro Levels, Level 4 - Macro Level and Level 5 - Mega Level.
9. *Five Level ROI Model.* One of the most challenging issues facing organizations is to measure the value on all types of projects and programs. The Return In Investment (ROI) methodology, developed by J. P. Phillips and P. P. Phillips [39], has proven to be an accurate, credible, and feasible approach addressing the accountability issues for all types of organizations. The methodology has been accepted and used in over 50 countries and is the leading approach to ROI accountability. The best approach to accomplish this balanced scorecard is the legendary and time-tested D. Kirkpatrick Model, with the additional fifth level added by P. Phillips. Level Five is not a D. Kirkpatrick step. D. Kirkpatrick alluded to ROI when he created level Four linking training results to business results.
10. *People-Process-Product Continuum Model.* B. Khan [24] developed a further evaluation model, the P₃ Model for e-learning. His main idea is evaluation should be included not only as last part of e-learning, but evaluation process must be embedded in each stage of e-learning: from planning until maintenance stages. Several phases of evaluation can be conducted during the overall e-learning process. These evaluations are conducted to improve the effectiveness of e-learning materials.
11. *PDPP Evaluation Model.* W. Zhang and L. Cheng [61] developed a four-phase evaluation model for e-learning courses, which includes planning evaluation, development evaluation, process evaluation, and product evaluation, called the PDPP evaluation model.
12. *Logic model for program Development and Evaluation.* The logic model for training program development and evaluation was created by

E. Taylor-Powell and E. Henert [51]. Their main idea is to use a logic model for planning and design of training programs.

13. *A Three-Layered Cyclic Model of E-Learning Development and Evaluation*. The e3Learning project: enrich, extend, evaluate learning (e3L) [31],
 - enrich learning
 - extend learning
 - evaluate learning

which was designed to assist teachers in web-assisted teaching and learning, conducted a wide range of e-learning evaluations. The e3L project operated across three universities, the Hong Kong Polytechnic University, the City University of Hong Kong, and The Chinese University of Hong Kong.

14. *The unfolding model*. In generalization of J. Messick's framework V. Ruhe and D. Zumbo [43] developed the unfolding model. This model can be adapted to different course designs and technologies. Evaluators can pick and choose from among the tools presented to tailor the evaluation to their own local needs. This feature is important because technology is constantly evolving. The paradigm shift from Web 1.0 to 2.0, in particular, has huge implications for evaluation practice. The power of this unfolding model is its adaptive and dynamic nature, which means that it can continue to be applied to new e-learning course designs from 'Web 1.0 and Learning Environments 1.0 to Web 2.0 and Learning Environments 2.0'.
15. *Fuzzy model*. This model introduces an evaluation methodology for the e-learners' behaviour which will be a feedback to the decision makers in e-learning system. Learner's profile plays a crucial role in the evaluation process to improve the e-learning process performance. The model focuses on the clustering of the e-learners based on their behaviour into specific categories that represent the learner's profiles. The learners' classes named as regular, workers, casual, bad, and absent. The model may answer the question of how to return bad students to be regular ones. This model has presented the use

of different fuzzy clustering techniques as fuzzy c-means and kernelized fuzzy c-means to find the learners' categories and predict their profiles. Fuzzy clustering reflects the learners' behaviour more than crisp clustering [20].

The above listed models were selected because they can be considered as representative for the recent evaluation models and practice. They form the basic models of actual evaluation theory. Of course there exist further models for special evaluation tasks. But we could not find an evaluation model or an example where clear and open is described how depending on the actual evaluation goal the collected data have to be processed in view of a quantitative evaluation.

Hence, an evaluation concept is needed allowing an evaluation of process success depending on the respective goal structure of process based on available observation data.

In the following sections we develop a corresponding concept. Starting with an analysis of logical goal structure of process to be evaluated we describe how goal structure adequate evaluation formulas and data processing rules can be obtained.

1.4 Summary

This section includes an abstract of thesis and takes reference to related work.

E-learning is becoming worldwide more and more in the focus of education systems. Leading countries like United States, European Union Countries, Canada, Australia and Korea, for instance, show a growing interest and increase their investments in e-learning. To measure the efficiency of these investments and to control and develop e-learning we need adapted evaluation technologies. Many researchers developed different approaches and models of evaluation. Essential contributions in this area were made by R. Tyler, D. Kirkpatrick, M. Scriven, M. Alkin, D. Stufflebeam, M. Patton, D. Ehlers, B. Khan, F. Colace, P. Lam, V. Ruhe and E. Taylor-Powell.

The second sub section surveys the basic ideas of evaluation research. E-learning is a core direction of a new era of education systems for each country. The local and global activities in e-learning are increasing all over

<i>Nr.</i>	<i>Evaluation models</i>	<i>Published</i>
1.	"Objective model"	1949
2.	"Kirkpatrick's four level of evaluation model"	1959
3.	"CIPP Model"	1960
4.	"Discrepancy Model"	1966
5.	"Connoisseurship Evaluation Models"	1969
6.	"CIRO model"	1970
7.	"Goal-Free Evaluation model"	1972
8.	"Kaufman's five Levels of Evaluation"	1995
9.	"Five Level ROI Model"	1997
10.	"People-Process-Product Continuum Model"	2004
11.	"PDPP Evaluation Model"	2007
12.	"Logic model for evaluation",	2008
13.	"A Three-Layered Cyclic Model",	2008
14.	"The unfolding model"	2008
15.	"Fuzzy model"	2010

Table 1: List of evaluation models.

the world. This development needs quality management and self assessment.

The challenge of the research ideas was to develop a new complete evaluation model for e-learning. The model should include all relevant components of an evaluation process. First and main part of evaluation process is planing part. We divided this part into three steps and visualize the e-learning goal by means of a logical structure. The logical structure of evaluation goal is one of new ideas of developed model. The next part of evaluation process is the data processing part. We developed based on general measure theory specific calculation rules for computation of scores by which it becomes possible to evaluate how the total goal of a logical oriented goal structure has been achieved. Data are collected by checklists which are adapted to the logical structure of evaluation goal.

In the third sub section we referred to L. Hogan, M. Scriven and M. Alkin's work. By our view in these works educational evaluation history and development is reviewed systematically and clear. With the "Historical de-

velopment review in evaluation 2007" L. Hogan [19] provided an overview on the historical evolution of program evaluation by describing seven significant time periods from the late 1700's through the 21st Century: The Age of Reform (1792-1900's), The Age of Efficiency and Testing (1900-1930), The Tylerian Age (1930-1945), The Tylerian Age (1930-1945), The Age of Development (1958-1972), The Age of Professionalization (1973-1983) and The Age of Expansion and Integration (1983-Present).

M. Scriven [47] defined three main revolutions in evaluation theory: The first revolution is to transform the status of evaluation from untouchable to respectable; The second revolution is to transform the status of evaluation from that of a respectable discipline to that of the alpha discipline and The third revolution is to extend the status of evaluation from that of alpha discipline to that of the paradigm applied discipline.

M. Alkin [1] first visualized Family Tree of evaluation theory. His book "Evaluation roots" includes an excellent analysis and review of whole evaluation theory and describes in detail the most important developments and approaches of evaluation theory.

2 E-learning framework and SURE model

2.1 The e-learning framework

E-learning is a complex process which includes different forms of educational technology. Each e-learning is supported by an e-learning framework. In the literature there are various definitions for e-learning frameworks. These definitions depend on the specific application background, learning and teaching methodology and technology.

L. Lucas formulated: *"e-Learning is the use of technology to enable people to learn any time and anywhere. e-Learning can include training, the delivery of just-in-time information and guidance from experts"* [32], D. Stockley [50] defines e-learning as: *"The delivery of a learning, training or education program by electronic means. E-learning involves the use of a computer or electronic device (e.g. a mobile phone) in some way to provide training, educational or learning material"*.

A more accepted definition for an e-learning framework has been presented by B. Khan [24]. His basic idea is to consider an e-learning framework as a multidimensional environment. B. Khan defines it by eight environments. By his definition an e-learning framework consists of the following environments:

"...The E-Learning Framework which can be used to capture an organization's inventory of e-learning by addressing issues encompassing the following eight dimensions of open and distributed learning environments:

- *Pedagogical: Refers to teaching and learning. This dimension addresses issues concerning content, audiences, goal and media analysis; design approach; organization and methods and strategies of e-learning environments.*
- *Technological: Examines issues of technology infrastructure in e-learning environments. This includes infrastructure planning, hardware and software.*
- *Interface Design: Refers to the overall look and feel of e-learning programs. The interface design dimension encompasses page and site design, content design, navigation, and usability testing.*

- *Evaluation: Includes both assessment of learners, and evaluation of the instruction and learning environment.*
- *Management: Refers to the maintenance of learning environment and distribution of information.*
- *Resource Support: Examines the online support and resources required to foster meaningful learning environments.*
- *Ethical: Relates to social and political influence, cultural diversity, bias, geographical diversity, learner diversity, information accessibility, etiquette, and the legal issues.*
- *Institutional: Issues of administrative affairs, academic affairs and student services related to e-learning.*

He further wrote:

...It is important to note that it may be difficult to come up with questions for some dimensions of the Framework for some specific stages of the e-learning phases which is understandable given the interrelationship among the characteristics of the eight dimensions..."

B. Khan's definition confirms that e-learning is a complex process.

Moreover, e-learning is a dynamical process which requires continuous improvement. Self evaluation of e-learning is one option for educators to upgrade e-learning step by step, part by part. According to B. Khan's definition evaluation is included like a further environment beside the other environments of framework.

By our understanding evaluation is not a separate entity of an e-learning framework. Evaluation is an essential component of any e-learning framework. We can use evaluation to identify and recognize good and weak points of e-learning. An in the e-learning environments embedded evaluation will support decision making during development of e-learning.

In this study we deal with a structure oriented evaluation model for the e-learning framework. In our model we use logical structures to visualize the structure of evaluation goal.

If we consider the e-learning framework of B. Khan, then the goals of e-learning evaluation are defined by the included environments. Each environment can be associated a corresponding key goal.

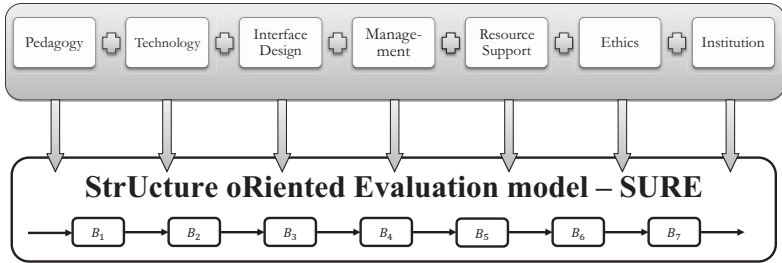


Figure 1: Evaluation and e-learning framework.

Figure 1 emphasizes our view to e-learning framework and evaluation. Our understanding of an e-learning framework differs from B. Khan's conception by the fact that we don't consider e-learning evaluation as a separated or additional environment of main framework. By our view evaluation has to be a part of each environment of e-learning framework.

As we discussed in Section 2.1 e-learning is a complex process. Different groups are involved into this process. For example: pedagogues, administrators, educators, multimedia designers, manager, stakeholders, decision makers, tutors, and last but not least the learners. Each included group has their own expectations on evaluation. What can be the expectations and understanding of these groups?

- Stakeholders and decision makers. They try to use e-learning evaluation as an instrument of quality control of the whole process. They also use it to assess the cost-efficiency of the teaching and learning process. Hence they expect a clear statistical analysis and measurement of success or progress of the e-learning process. Therefore it is sometimes more important for them to get a confirmation of correctness of their decisions in the frame of an investment project than further hints for the further development of e-learning. For them evaluation is decision support for the economical management of corresponding projects. Thus the stakeholders and decision makers are indirectly involved in the e-learning process, they look to that from a distance and

observe how it is running. In this sense, for them evaluation is a control and monitoring tool for e-learning.

- Lecturers, teachers and tutors. They are the main group which is deeply involved in the e-learning process. They play the major role at content design. Tutors lead and implement e-learning. They are interested in corresponding feedback for further improvement and development of their teaching. Their focus is not only directed towards evaluation and quality measurement, but to get suggestions about weaknesses and opportunities for improvement of e-learning. Their central focus is on designing quality, content and learning environment for the learners.
- Multimedia designer and technical support center. This group has to fulfill a responsible task in the e-learning process. Without the corresponding technological environment e-learning loses its meaning. In this context evaluation is used to find out what the learners, the tutors and other groups expect from the technical support center. For instance, are all parties satisfied with technical support service, are there any weak points or enough reserves in the technical environment, correspond user interface and added value aspects in the e-content the expectations of learners and tutors, have the multimedia designers developed user friendly human-computer interaction interfaces for e-learning environment etc. For multimedia designer the added value effect to e-content is significant. Hence the highlighted focus for them is the virtual environment which has to support learners and tutors during e-learning.
- Administrators and managers. Main task of this group is to manage the whole administration processes which ensure successful e-learning. This includes introduction of announcements about courses, registration fee transactions, midterm and final exam process and finally certification and evaluation of whole e-learning process. Success of the administration process depends essentially on learning management system of e-learning. A well designed and programmed learning management system includes all necessary administration modules and tools. For administrators and managers the highlighted focus is to know whether the learners have faced any problem with

regard to registration, whether the information about courses helped them, how successfully the e-learning process runs, and how satisfied are the learners.

- Learners and students. Learners of e-learning can be different target groups. These groups can be different depending on the focus of the educational institutions. They can be students of universities or colleges, adult people or children from elementary and high schools. Irrespective of the target group, if the learners who are attending to the e-course can not accept or enjoy the learning process, e-learning will not run successfully. Although the e-learning environment is up to high standard and the tutors are professional, if learners are not cooperative enough, then the learning process of e-learning will not run successfully. To motivate learners and make them active during study is one of the main challenges of e-learning. To improve this part, educational institutions have to identify needs and expectations of learners. To get feedback from students surveys or questionnaires can be used. Satisfaction of learners is always one of the main aim of educational evaluation.

2.2 The need of evaluation in e-learning process

E-learning is result of teamwork into which different groups, experts or partners are involved. Two main groups are of particular importance. On the one hand the e-learning developers, and on the other hand the users like students or learners which often have contradictory expectations and interests. Usually the financial budget of an e-learning project is limited. Educators who develop e-learning have to negotiate how to equalize all interests under the given financial budget. An additional barrier in that situation is that the creation and implementation of an e-learning is frequently more expensive than the preparation of a traditional face to face teaching and learning.

All these aspects confirm that educational institutions have to take into consideration different and sometimes contrary interests during the development of e-learning. A solution for this situation is an evaluation model which covers the interests of all involved groups in an e-learning. Such an evaluation model takes reference to quality and other aspects of involved

groups. This evaluation model has to reflect and process all dimensions of involved aspects and criteria in a consistent manner. Moreover, the evaluation model should be able to take reference to different expectations of involved groups in the respective process. These groups are,

- Educational institutions
- Faculties
- Professors
- Lifelong learners
- Stakeholders
- Companies
- Decision makers
- Financial Backers etc.

A similar approach is also suggested in the European Commission handbook for e-learning evaluation [5]. In Section 1 they write:

"...The development of e-learning products and the provision of e-learning opportunities is one of the most rapidly expanding areas of education and training. Whether this is through an Intranet, the internet, multimedia, interactive TV or computer based training, the growth of e-learning is accelerating. However, what is known about these innovative approaches to training has been limited by the shortage of scientifically credible evaluation. Is e-learning effective? In what contexts? For what groups of learners? How do different learners respond? Are there marked differences between different Information and communication technology platforms? Does the social-cultural environment make a difference? Considering the costs of implementing Information and communication technology based training, is there a positive return on investment? What are the perceptions of VET professionals? What problems has it created for them? E-learning is also one of the areas that attracts the most research and development funding. If this investment is

to be maximized, it is imperative that we generate robust models for the evaluation of e-learning and tools which are flexible in use but consistent in results.

'Although recent attention has increased e-learning evaluation, the current research base for evaluating e-learning is inadequate Due to the initial cost of implementing e-learning programs, it is important to conduct evaluation studies' (American Society for Training and Development, 2001)..."

The number of e-courses and online courses is increasing rapidly.

<i>e-course</i>	<i>online course</i>	<i>Date</i>
507 000 000	293 000 000	November, 2012
1 360 000 000	1 190 000 000	August, 2013
1 400 000 000	1 250 000 000	October, 2013
1 510 000 000	1 290 000 000	November, 2013
1 520 000 000	1 310 000 000	December, 2013
1 670 000 000	1 520 000 000	January, 2014

Table 2: Google search engine result.

Table 2 shows a result of a Google search for contents in the internet with keyword "e-course" and "online course". These facts confirm once more the need of a corresponding quality evaluation information as orientation for learners who want to study any topic based on e-learning. This holds analogously for financial backers who are ready to support e-learning projects.

Finally, educational institutions and interested groups need evaluation and monitoring progress of actual e-learning projects.

A core problem of existing linear models for e-learning evaluation is that clearly defined weight factors do not exist for evaluation of different groups which are involved in an e-learning. Each group can attest that their part is more important than others. Such conflicts can be avoided by structure oriented evaluation.

By the structure oriented evaluation model the interests of different groups can be mapped as key goals with a special logical inner structure.

Then it can be evaluated how each group has achieved its key goal. Herein, a special evaluation score can be assigned to each group via a corresponding checklist results. Moreover, by the product rule (see Section 3), a consistent evaluation of whole process becomes possible as a general evaluation of an e-learning process. The depth and fineness of included logical structures for the key goals is very variable and adjustable.

The reasons for the need of evaluation for e-learning processes:

- Measurement of success and failure of e-learning process. Only evaluation can show how an e-learning process is running. Without any evaluation one can't discuss about success or failure of e-learning. Evaluation results make it possible.
- Cooperation of involved groups at the same level in the e-learning. E-learning includes different groups with distinguished tasks. All involved groups have to cooperate to design and implement successful e-learning. The evaluation result can show all involved groups how successful their collaboration has been during the development and implementation period of the e-learning process.
- Recognizing as early as possible deficiencies of a running e-learning process. To recognize weaknesses of e-learning is important to design and implement corresponding instruments for quality control.
- Improving e-learning course quality. The quality is the mission of each e-learning process. By structure oriented evaluation model educational institutions can assess quality of e-learning more objectively and transparently. That will help the interested groups to accept the results of evaluation more confidently.
- Control the running e-learning. Complex processes need control. The embedded formative evaluation can help the educators to control ongoing e-learning processes.
- Support of decision making. E-learning is expensive. To make decisions on e-learning project, stakeholders need formative and summative, pre or post evaluation results. These results can be helpful for better decisions making.

- Improvement of e-learning as a whole. Each e-learning process consists of several environments. In order to provide successful e-learning all environments should be supported to run effectively. An embedded process evaluation can show how e-learning can be improved in detail and as a whole.
- Preparation for international accreditation of e-learning. Local evaluation activities can support the e-learning process to be prepared for international accreditation. Evaluation of own e-learning courses by inner evaluation can show the educational institutions how a course can be improved in preparation for an international accreditation.

2.3 The structure oriented evaluation model

"... When the needs and expectations for a program or process are defined, developing specific, measurable objectives that will achieve these expectations becomes easier and increases the likelihood of program success". "... If the needs important to all stakeholders are not clarified in the beginning, then measures taken during the evaluation are based on guesswork..." [41].

Patricia Pulliam Phillips (2010). (President and CEO of the ROI Institute, an expert in measurement and evaluation).

2.3.1 Basic idea of structure oriented evaluation

Depending on a special e-learning unit and according to the visions of included groups, the goal of evaluation can be different for each evaluation team. An evaluation team has to include members as experts or agents from all involved groups in e-learning. The evaluation team has to take into account the special goals of all groups.

The basic idea of our evaluation model is that each reasonably formulated evaluation goal has or should have a logical inner structure. The challenge of this research was to develop a transparent evaluation model which

takes into account the logical structure of evaluation goal with the side condition, that all steps of evaluation are clear, understandable and traceable for the evaluation team. The first step in the application of this model is always the analysis and definition of logical structure of evaluation goals. Under this aspect, we named the model as *StrUcture oRiented Evaluation* model (SURE model).

The theoretical basis of SURE model forms the general measure theory. The idea is that, if one wants to evaluate or measure how an e-learning has achieved its goal, that should be done in line with measure theoretical rules and principles as they are used in proven manner in geometrical context at measurement of lengths, areas or volumes, in natural sciences and in probability theory, for instance.

In the SURE model, the evaluation goal is understood as a collection of partial or sub goals which are embedded over key goals into a main goal. All sub goals contributes to the success of main goal. Between sub goals consist logical relations which describe how the sub goals have influence to achieving of the main goal.

The partial goals can be divided into two groups of goals: key and sub goals (see Figure 2). The key goals have to be achieved in order to the main goal is achieved. The sub goals describe possible alternative goals for the achievement of a key goal. Sub goals can, but must not associated with key goals. It depends on the structure and complexity of main goal of evaluation.

The evaluation by SURE model begins with the definition of key goals of evaluation. That is written sometimes in literature like: We have "to define needs of stakeholders". In our model evaluation team has to define the "needs of involved groups". This is a difficult step at the application of each evaluation model. The evaluation team has to define the key goals of evaluation in accordance with the mission of evaluation. The main questions in this context are: What we want to evaluate? What the stakeholders want to measure by the evaluation process? What the stakeholders expect from evaluation results? The selection of corresponding tools for this process is a matter of evaluation team. It is important to specify all needs and expectations of stakeholders and involved groups on the evaluation process.

By measure theory it becomes possible to include all goals into one common product space and the SURE model can calculate a global evaluation

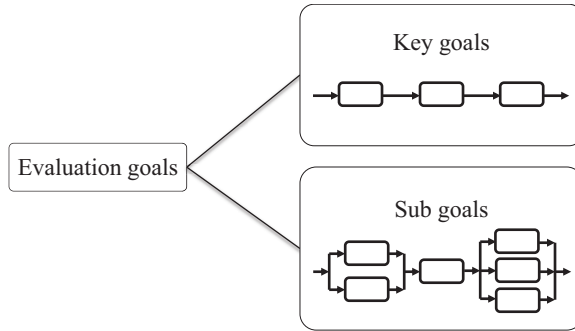


Figure 2: Types and visualization of evaluation goals in SURE model.

value. Moreover, if it is necessary evaluation team can define embedded or levelled logical structures of the evaluation goal. All further steps follow the same method for data collection and data processing.

The structure of the evaluation goal can be different depending on the interests and expectations of the evaluation team. Figure 3 shows two different types of key goal structures. Figure 3a emphasizes that depending on status of stakeholders, key goals can have different levels. Each level can be evaluated via adapted key goal structures. Figure 3b is visualizing the embedded key goal structures of different groups with specific tasks. In this case level can be the same but expectations can be different. That means, each interested group can have an own key goal structure.

During the definition of evaluation goal, the graphical visualization of logical structure of evaluation goal is helpfully. Such a visualization supports the analysis of main evaluation goal and forms a basis for creation of an adapted checklist.

For example, to evaluate an e-learning course we have different ways:

First way: All aspects of e-learning are embedded into one main objective as key goals. If even one of these key goals is failed then evaluation result should show that the goal of e-learning has been failed. This is a strong condition. Because each part of e-learning depends on the success of

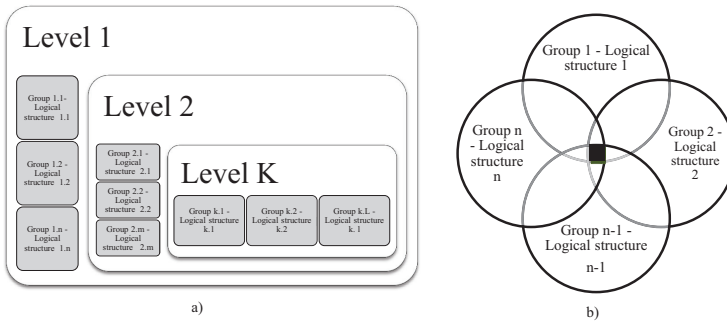


Figure 3: The different types of logical structures.

others. But their tasks are different. For example, tutors work well, and it has been confirmed by learners by top scores. But if the level of satisfaction of learners about the learning management system was not high enough and if they have evaluated that as bad, the effort of tutors becomes invisible. The reason is, that all key goals are embedded as a series scheme into the main goal.

This simultaneous inclusion of all groups into a key goal structure can cause several side effects. To avoid such effects, and for a more fair evaluation it is better to create distinguished logical structures of key goals which are adapted according to the interest and expectations of different groups.

Second way: Evaluation team can consider different logical structures for each involved group. For instance, a logical structure of key goals can be created assuring the content quality which matches with the expectations of content designers. As the content quality depends on many factors, evaluation team can identify the main factors which can be included in key goal structure. A similar structure can be adopted to evaluate tutors too. Their skills can be evaluated using other criteria than content quality, and the evaluation team can define these factors. This is emphasized in Figure 3. The Figure 3a shows that decision makers of different levels can have their own logical structures of evaluation goals such that we can obtain a levelled logical structure. Figure 3b gives another way of creating the logical structures in a similar manner. Here each part of diagram defines a

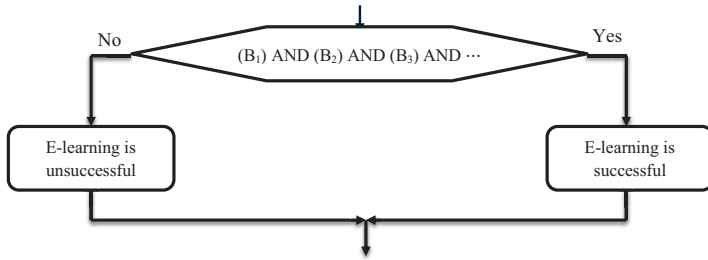


Figure 4: Algorithm for logical structure of key goals.

logical structures related to a particular group. The middle area covers the common interests of all involved groups in e-learning.

At the very onset of definition of key goals, evaluation team has to decide which element of the process will be evaluated by the corresponding logical structure of evaluation model. For example this can be: complete e-learning program, e-learning project, e-learning course, learning management system, teaching faculty, administration part, content quality, technical support for e-learning, cost of investment for e-learning project, etc. Depending on the evaluation object or the process key goals of logical structures can be differently embedded into other key goals.

During planning and designing evaluation team always has to focus on the correct definition of evaluation goals. How the key and sub goals are to define by evaluation team this will be discussed later. First we will explain what we want to understand under key and sub goals.

Key goals are goals which are linked by logical 'AND' and can be visualized by a logical series structure. Order of key goal has no meaning. In this sense the key goal B_1 has same 'value' for system like any other key goal $B_i, i = 2, \dots, r$.

Formal we can collect this together as follows:

IF (B_1) AND (B_2) AND (B_3) AND ... AND (B_n)
 THEN *E-learning process is successful.*

This pseudo-code emphasizes the calculation algorithm of the example in Figure 4. The calculation algorithm proves that the evaluation result is successful, only if all key goals are successful. Hence the AND condition as it is used in programming too is a very strong condition. Irrespective of number of key goals the total goal is achieved when all elements reach their goals. That means, each involved group can have their own key goals in the logical main structure. For example: The quality of content design, multimedia environment, tutors, administration part or technological support can be considered under the key goals in SURE model. If each single key goal is evaluated as successfully, the total result of evaluation will be successfully too. This indicates that collaboration, cooperation and team work can contribute to success of e-learning processes.

The logical structure of evaluation goal confirms not only how important collaboration is in e-learning but it evaluates the success of each key goal. By this the involved groups of e-learning can monitor and recognize good and weak components of the whole process.

Key goals can be considered as the main dimensions of evaluation. As a rule these goals can be achieved on different ways or by different methods. In this case a key goal can be characterized further by sub goals. The sub goals of a key goal are connected by logical OR and can be emphasized by a logical parallel structure.

That means, a key goal is achieved then if at least one of its embedded sub goals is achieved. For example, if e-learning material is delivered using different media: reading materials, audio version and video version, learners can use all these possibilities parallel or can select one or two of them. All versions have the same content. In this case evaluation team can use a parallel structure for e-learning content design. If learners evaluate just one of them successfully the main key goal is evaluated as successfully.

This is expressed by the following pseudo code: If B is a key goal which consists of the sub goals A_1, \dots, A_s then we have

IF (A_1) OR (A_2) OR (A_3) OR ... OR (A_s)

THEN

Key goal B has been achieved.

The number of sub goals of key goal should be kept small. The reason

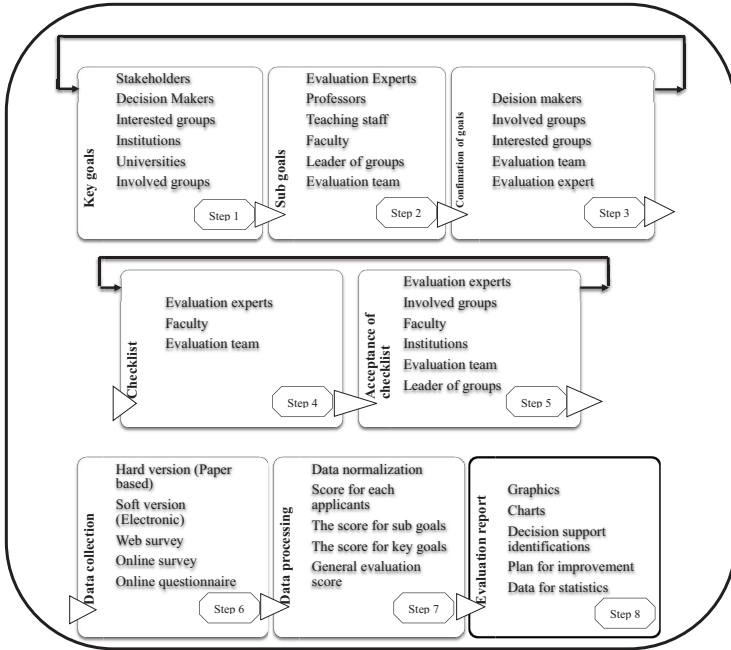


Figure 5: The structure oriented evaluation model.

is that in other cases already small score values of sub goals can yield to an unreasonably high score of a corresponding key goal. In such cases one should modify the logical structure for instance by defining of further or finer key goals.

2.3.2 The eight steps of SURE model

The SURE model consists of eight steps, see Figure 5. All steps of SURE model have a specific meaning. Output of the previous step will become the input of next step.

That are the eight steps of SURE model:

- Definition of key goals
- Definition of sub goals
- Confirmation of evaluation goals
- Creation of checklist
- Acceptance of checklist
- Data collection
- Data processing
- The evaluation report

In the following we will explain the single steps of SURE model.

Step 1. Definition of key goals

In the first step of SURE model, evaluation team has to define the key goals of evaluation. These goals are essential for main mission of evaluation.

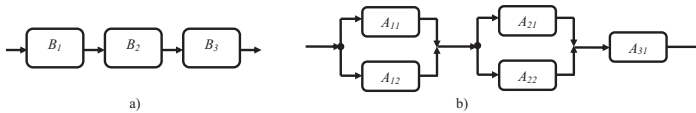


Figure 6: The logical structures of evaluation goals.

For example, we have to work on an actual running e-learning process. In this case we start with definition of key goals. These could be, for instance: Registration process (B_1), Course material (B_2) and Tutor skills (B_3). These key goals can be visualized by a logical series structure (see Figure 6a). These components were selected as key goals because without any registration learners have no right to access the e-learning framework. After successful registration learners can start the learning process. Without any e-content or the learning materials an e-learning cannot run. If the learners

are studying the e-materials without any tutoring, learners may lose their learning interest in e-learning and they can drop out from the study. The learners need tutor's support and feedback during learning. All these facts show that these goals could be considered as key goals of evaluation.

Successful evaluation is then a result of:

- Administration
- Course material
- Tutor skills

If only one of these goals is failed, then the e-learning process will be evaluated as failed.

Step 2. Definition of sub goals.

During the second step, the key goals defined in the first step, have to be described in detailed manner if necessary. This can be done by sub goals. Sub goals can be understood as different ways to reach a key goal.

Example 1: To evaluate the performance of a registration process we have two possibilities or ways as a rule: full online or blended registration. The corresponding key goal includes then two sub goals A_{11} and A_{12} which refer to online and blended registration. The next key goal B_2 concerns the course materials, for instance: this can be further divided into sub goals which refer to reading material quality (A_{22}) and audio and video material quality (A_{22}). Course content can be designed in different ways: reading materials which can be downloaded or read online; audio records which can be listened online; video materials which are included into course learning materials. Key goal B_3 could be focused on Tutor. In this case the key goal B_3 is defined via a single sub goal A_{31} .

Example 2: The e-learning process needs tutors. Tutor's main task is to support learners during e-learning process. But a tutor system is not an example for a parallel structure in the logical structure of an evaluation goal. An other example for parallel observation of an evaluation goal is considered Section 4.4.

Key goal	Sub goal	Question
B_1	A_{11}	Online registration process was easy to use
	A_{12}	Blended version of registration process was well organized
B_2	A_{21}	Course content quality was high
	A_{22}	Course material level was high
B_3	A_{31}	Tutors support was very useful

Table 3: Checklist design proposal.

Step 3. Confirmation of evaluation goals.

The evaluation team has to confirm the final version of evaluation goals which have been defined during the previous two steps. If necessary, evaluation team can define embedded or levelled logical structures until the evaluation goals are fully accepted by all members of the evaluation team. Further, the checklist for survey has to be adapted to the accepted logical structure. When all members of evaluation team agreeing to logical structure of evaluation goals, that should be fixed by a protocol. This is helpful in avoiding conflicts between evaluation team and stakeholders.

Step 4. Creation of checklist.

The checklist is a well developed data collection method. There exist many software solutions to create a corresponding online survey. For example: Monkey [99], fluidsurveys [100], iPerceptions [101], free online survey [102], kwik survey [103], easy polls [104], survey planet [105], Sogo survey [106], eSurveypro [107], esurvey creator [108], Stellarsurvey [109], Questionpro [110], esurv [111], questionform [112], panel place [113], survey crest [114], addpoll [115] and Quick serveys [116].

For further discussion on this issue we refer to Penny [37], Betsy [7], Paul [8], Blayney [9] and Bridge [10]. The checklist of SURE model has to be created based on the sub goals of logical structure. Table 3 shows a proposal for checklist design for Examples. The generation of checklist can be supported by corresponding software.

However, existing software solutions cannot take reference to logical structure of the evaluation goal of SURE model. For the application of

SURE model we need a corresponding implementation tool. This implementation should include functions like: checklist generation, data collection and data processing.

Step 5. Acceptance of checklist.

Only an accepted checklist should be used for data collection. Clear formulation of questions is only one aspect of checklist. A further important aspect is the design of checklist. Each member of the evaluation team has to check that before confirmation and, if necessary, appearance and design of checklist has to be improved.

Step 6. Data collection.

There are several techniques for data collection: surveys and questionnaire, tests and assessments, interviews, focus groups, action plans, case studies, and performance records [41]. Evaluation team can use any of these techniques. However, for the SURE model an online survey is recommended. There is a beta version tool for SURE model (see Section 2.3). Via online survey data can be tabulated and processed automatically. This prevents human errors that usually occur during transferring the collected data to the data sheet.

Step 7. Data processing.

These are the main steps of data processing: Let us consider an evaluation structure C which consists of r key goals B_i , $i = 1, \dots, r$, and each key goal consists of s_i sub goals A_{ij} , $j = 1, \dots, s_i$, $i = 1, \dots, r$. Then we have

$$C = \bigcap_{i=1}^r B_i = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}.$$

We suppose that we have n checklist results obtained by an checklist adapted to goal structure C . The evaluation interval for sub goal A_{ij} be the interval $[x'_{ij}, x''_{ij}]$. Let

$$\begin{aligned} & x_{11}^{(1)}, \dots, x_{1s_1}^{(1)}, x_{21}^{(1)}, \dots, x_{2s_2}^{(1)}, \dots, x_{r1}^{(1)}, \dots, x_{rs_r}^{(1)}, \\ & \dots \\ & x_{11}^{(n)}, \dots, x_{1s_1}^{(n)}, x_{21}^{(n)}, \dots, x_{2s_2}^{(n)}, \dots, x_{r1}^{(n)}, \dots, x_{rs_r}^{(n)}, \end{aligned}$$

be the corresponding sampling results. Here denotes $x_{ij}^{(k)}$ the obtained an-

swer of k^{th} student to checklist question how the sub goal A_{ij} has been achieved.

Then the empirical score (computed score based on sampling results) that the aim of goal structure C has been achieved is calculated by

$$Q^*(C) = \frac{1}{n} \sum_{k=1}^n Q^{*(k)}(C) = \frac{1}{n} \sum_{k=1}^n \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)}) \right). \quad (2)$$

Here denotes $q_{ij}^{*(k)}$ the empirical score for sub goal A_{ij} according the k^{th} checklist result, $k = 1, \dots, n$. It is obtained by normalisation of checklist value $x_{ij}^{(k)}$. It holds

$$q_{ij}^{*(k)} = \frac{x_{ij}^{(k)} - x'_{ij}}{x''_{ij} - x'_{ij}}.$$

Beside the empirical score $Q^*(C)$ we can get estimation values for the key goal scores $Q(B_1), \dots, Q(B_r)$ as well as for the sub goal scores $Q(A_{11}), \dots, Q(A_{rs_r})$. These are special cases in sense of formula (2). We get

$$Q^*(B_i) = \frac{1}{n} \sum_{k=1}^n \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)}) \right)$$

and

$$Q^*(A_{ij}) = \frac{1}{n} \sum_{k=1}^n q_{ij}^{*(k)}.$$

The score values $Q^*(C)$ and $Q^*(B_i)$ are yet to transform by calibration into the final empirical evaluation scores $Q_e^*(C)$ and $Q_e^*(B_i)$. It holds

$$Q_e^*(C) = \frac{1}{n} \sum_{k=1}^n \sqrt[r]{\prod_{i=1}^r \left(1 - \sqrt[s_i]{\prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)})} \right)} \quad (3)$$

and

$$Q_e^*(B_i) = \frac{1}{n} \sum_{k=1}^n \left(1 - \sqrt[s_i]{\prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)})} \right) \quad (4)$$

as well as

$$Q_e^*(A_{ij}) = Q^*(A_{ij}). \quad (5)$$

The evaluation scores allow a comparison of score values between different goal structures. The empirical evaluation score can be interpreted as an index of satisfaction of learners with an e-learning. A value $Q_e^*(Q) = 0.5$ reflects an average level of satisfaction. Values over 0.5 an above average, values less than 0.5 a below-average satisfaction. The precision of obtained estimation values can be described by confidence intervals. For details we refer to Section 3.

Step 8. The evaluation report

By SURE model quantitative and qualitative results can be obtained. The outcome of first five steps forms the qualitative part of evaluation, steps 6 and 7 the quantitative part. The evaluation report has to include both parts of the evaluation process. The evaluation report has to make visible outcomes and findings of the evaluation process.

The kind of reports can be different depending on the audience to whom the report is addressed. By SURE model we can calculate several evaluation scores. Which evaluation score should be used for what, is to decide by the evaluation team. The graphics and charts are generated by SURE model application automatically, and these graphical representations can be used for the report.

Evaluation report has to be delivered to stakeholders or audiences at appropriate times. Else all effort of evaluation team could have been in vain. As a rule all interested groups expect a quick report after data collection.

2.4 The tool for SURE model

In frame of our research, we developed a beta version of an online tool for SURE model.

Figure 7 shows waterfall model of tool for SURE model. The main architecture consists of five units.

1. The logical structure of evaluation goals. This unit includes functions to create the key and sub goal sets and to fix the logical structure. This structure is used in the fourth unit of architecture for data processing.

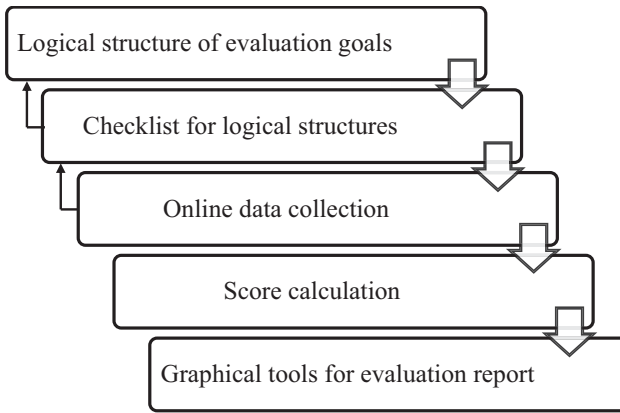


Figure 7: The waterfall diagram for SURE model.

Here we have to include confirmed evaluation goals. The results of first 3 steps (see Figure 5) will input of this step.

2. The checklist for logical structure. Here user can enter checklist questions using the checklist tool. Result of this unit will be an online survey questionnaire for SURE model. Here we will use results of step 4 of SURE model.
3. Online data collection. The function of this unit is to support users during online survey. Admin of tool can control and check data collection process via this unit (Step 6 of SURE model).
4. Automatic processing of collected data. One of the advantages of SURE model are its adapted calculation rules for data processing (see Section 3). All collected data are tabulated and will be processed by the SURE model calculation rules automatically (Step 7 of SURE model).
5. Graphical tools for evaluation report. This beta tool for SURE model includes simple functions to support evaluation report by evaluation

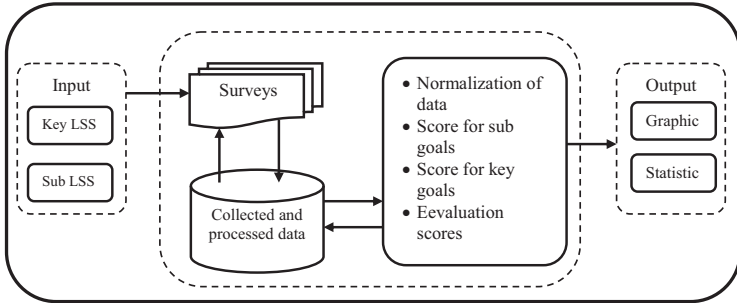


Figure 8: The architecture of tool for SURE model.

scores. Evaluation team can produce different types of charts and graphics with this unit (Step 8 of SURE model).

Figure 8 shows the software architecture of SURE model tool. Its architecture corresponds the waterfall diagram. Input of tool is the logical structure of evaluation goal. The SURE model can save several surveys in a Survey library. All data are collected in a database. The database consists of collected original data and calculation results. Data processing part computes evaluation scores according to the calculation rules which are presented in detail in Section 3. The output unit can generate graphics and statistics for empirical evaluation scores.

The access to tool distinguishes between two types of users: 'Admin' and 'User'. Admin is a user who has access to all five units of tool. Admin has following rights:

- Create new surveys
- Edit existing surveys
- Manage comment fields
- Generate accounts for users
- Check survey process

- Create sheets for accounts
- Tabulate collected data
- Process collected data
- Tabulate computed scores
- Create graphics by scores etc.

User is a person who receives an unique access code from the Admin. User has overview on answer progress. User can submit answers only if all questions are answered.

Accordingly users have following possibilities:

- Access to open survey using own code
- Interrupt answer and continue later
- See statistic about own answer progress
- Change answer before submission
- Insertion of comments to questions
- Submit completely answered survey

2.5 Summary

In this section we introduce the basic ideas of structure oriented evaluation model (SURE model). The SURE model developed for evaluation of e-learning takes reference to the logical structure of evaluation goals. Main target of SURE model is to create a complete evaluation model for e-learning which has qualitative and quantitative properties. The model should be easy to implement and control. Traceable and transparent calculation rules for data processing are helpful to evaluation team to process collected data objectively.

Section 2.1: This section describes the main concepts of the e-learning framework based on the definition of B.Khan [24]. B.Khan classified the

e-learning process in eight dimensions of open and distributed learning environments: pedagogical; interface design; evaluation; management; ethics; resources and support, institutions and technological. Multidimensional environment of e-learning framework proves that e-learning is a complex system where different groups have to cooperate together. Difference between our view to e-learning framework in comparison with B.Khan is that the evaluation environment has to be included in each single environment of an e-learning framework (see Figure 1). By our opinion evaluation process has to cover each environment of e-learning framework. Moreover, the need of evaluation in e-learning is discussed. In the e-learning process different groups are involved with specific tasks. We discuss, which groups are interested in evaluation of e-learning and what could be expected from evaluation. The different interests of involved groups into an evaluation process can be taken into account via adapted logical structures of evaluation goals.

Section 2.2: This section is focused on basic idea of the developed evaluation model. The structure oriented evaluation model consists of eight steps. Here we explain in detail function of each step. In step 1, evaluation team has to define key goals of evaluation. Second step is for definition of sub goals. In third step evaluation team has to confirm logical structure of evaluation goals. If a general consensus on evaluation goals has not yet achieved all members of evaluation team have to repeat first two steps until consensus is reached. After confirmation of logical structure of evaluation goals, in fourth step, the evaluation team has to create the checklist. In next step evaluation team has to confirm the final checklist. In seventh step collected data are processed by adapted calculation rules of evaluation model. In the last step the evaluation team prepares an evaluation reports.

Section 2.3: In this section we present a short introduction into a beta version tool of SURE model. The tool consists of five units. Each unit is related to steps of evaluation model. The tool supports a structure oriented evaluation in sense of SURE model. It is focused on evaluation team and users. The evaluation team can use the tool with Admin right and can create online surveys, organize data collection, process data and prepare an evaluation report based on calculated evaluation scores.

3 Theoretical foundation of the SURE model

In this section we consider the theoretical base for data processing step of SURE model. As explained before, each e-learning consists of different dimensions. Hence for evaluation of e-learning we need a multidimensional evaluation method which, moreover, is able to take reference to co-operation structure of involved components. Furthermore, an evaluation method should satisfy the same basic rules as they hold for measurement procedures in geometry, integration theory or probability theory, for instance.

The measures used in these fields are special cases as they are considered in general measure theory. Today measure theory is the generally accepted theoretical background for measurement and evaluation of most different objects. That means, one should use this theory too at evaluation of e-learning.

Under this aspect we develop a measure theoretical data processing method for evaluation of e-learning. We start with a short analysis of an e-learning process. After that we describe the evaluation of involved process components by measure spaces. These spaces form the base for a corresponding product measure space. With this product space a multidimensional structure oriented evaluation of an e-learning becomes possible.

Depending on how the involved groups cooperate together to reach the e-learning goal we distinguish two basic types of goal structures of an e-learning. These are the series and parallel structures. For these goal structures corresponding score calculation rules are derived. By combination of these rules we can calculate moreover scores for evaluation of more complex e-learning.

For calculation of scores of an e-learning we need the scores of involved process components. We describe a corresponding estimation method for the score of an e-learning based on sampling results obtained with an adapted checklist.

To make scores for different e-learnings comparable we present a score calibration procedure. By this procedure the score for an e-learning is transformed into a so-called evaluation score. For the details and properties of this transformation we refer to Section 3.4. This evaluation score opens fur-

ther interpretation possibilities which are helpful for analysis of evaluation results.

Based on checklist data we get an empirical evaluation score. The precision of empirical evaluation scores can be estimated by confidence intervals.

In last subsection we consider some simulation examples. These examples show how the presented evaluation approach works in typical evaluation situations.

3.1 Logical structure and evaluation model

E-learning is a complex process which consists of several environments. To evaluate the efficiency of a given e-learning process we have to define what we want to understand formally by an e-learning as 'process'.

Each process consists of several process components. They are as partial processes involved into the achievement of overall process goal. This leads to the following definition.

Definition. Let $\mathcal{G} = \{A_1, \dots, A_m\}$ be a given set of process components where to each component a functional goal is associated. A sub set $\mathcal{B} = \{A_1, \dots, A_r\} \subseteq \mathcal{G}$ of process components which interact in sense of the achievement of a defined overall goal is said to be a *process*.

Hence a simple random selection of process components forms as a rule not yet a process. The process components have to work together with respect to a defined overall goal. That means, beside the process components a process is characterised by structural properties. These structural properties describe how the involved process components contribute to the achievement of total goal.

The functional structure of a process can be characterised formally by logical operations like AND and OR. In set-theoretical interpretation these are intersection and union. The process components of a process can interact together in different manner. Depending on how the process components of process contribute to achievement of total goal we distinguish two basic process structures.

Definition. A process $\mathcal{B} = \{A_1, \dots, A_r\} \subseteq \mathcal{G}$ is a *series process* if the goal of process is achieved if and only if each process component A_1, \dots, A_r

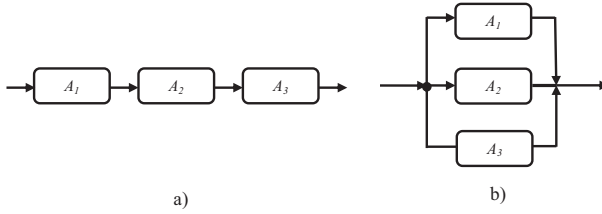


Figure 9: Series and parallel process.

achieves its goal.

This property of a process can be described formally as follows. Let $\mathcal{B} = \{A_1, \dots, A_r\} \subseteq \mathcal{G}$ be a series process consisting of r process components A_1, \dots, A_r . Denote by B and A_1, \dots, A_r the fact that the process \mathcal{B} and the process components A_1, \dots, A_r achieve their goals, respectively. Then for a serial process the relation between B and A_1, \dots, A_r can be formally represented by

$$B = A_1 \cap \dots \cap A_r = \bigcap_{i=1}^r A_i.$$

That means, a series process \mathcal{B} achieves its total goal if all included process components achieve their goal. This can be emphasised graphically by a series scheme as it is shown in Figure 9a. The order of process components in a series scheme is without of significance.

Beside series processes the following process type is of particular relevance.

Definition. A process $\mathcal{B} = \{A_1, \dots, A_s\} \subseteq \mathcal{G}$ is a *parallel process* if the goal of process \mathcal{B} is achieved if at least one of the process components A_1, \dots, A_s achieves its goal.

Formally then holds

$$B = A_1 \cup A_2 \cup \dots \cup A_s = \bigcup_{i=1}^s A_i.$$

This can be emphasised graphically by a parallel scheme like in Figure 9b. The total goal of a parallel process is already achieved if one of the process components has reached its goal.

By combination of series and parallel processes more complex process structures can be obtained.

Two processes \mathcal{B}_1 and \mathcal{B}_2 can overlap. In this context we define yet the following.

Definition. Two processes $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{G}$ are said to be *separate*, if

$$\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$$

holds.

Now we will describe how we can evaluate a process depending on its logical goal structure. For that we consider a process which consists of k , $k \geq 1$, process components A_1, \dots, A_k , which cooperate together in view of achieving of a defined process goal. Our aim is to evaluate in which quality the process has achieved its process goal.

For that we develop a method for evaluation of e-learning by means of general measure theory. Our aim is an evaluation approach whose score calculation rules are compatible to the corresponding rules as they are used in established manner in geometrical context at measurement of length, areas or volumes or in stochastics at calculation of probabilities or reliabilities, for instance. By means of measure theory it becomes possible to evaluate the performance of an e-learning on the same theoretical base as it is done on the fields mentioned above. Subjective influences at evaluation like one has that for instance at the use of weight factors are excluded then.

For the measure theoretical description of evaluation of process quality we have to proceed in two steps: First we have to describe the evaluation of process components A_1, \dots, A_k in measure theoretical sense by adapted measure spaces. After that we can join together these spaces to a common product space by which the evaluation of process performance with respect to a given process goal structure becomes possible.

I. Measure spaces for the process components. For the evaluation how an process component $A_i \in \{A_1, \dots, A_k\}$ has achieved its process goal we consider a measure space $(\Omega_i, \mathcal{A}_i, Q_i)$. A measure space consists in sense of general measure theory of three objects Ω_i, \mathcal{A}_i and Q_i which can be defined here as follows:

1. Let $\Omega_i = \{\omega_{i1}, \omega_{i2}\}$ be a two-element set. The elements of this set we

denote as *elementary goals*. In this sense the element ω_{i1} is standing for: the process goal of process component A_i has been achieved (is achieved), the element ω_{i2} is standing for: the process goal of process component A_i has not been achieved (is not achieved).

The set Ω_i is denoted as *goal space* of process component A_i .

2. Let \mathcal{A}_i be the set of all subsets of goal space Ω_i . Then we have

$$\mathcal{A}_i = \{\emptyset, A_{i1}, A_{i2}, \Omega_i\},$$

The sub sets A_{i1} and A_{i2} are defined by $A_{i1} = \{\omega_{i1}\}$ and $A_{i2} = \{\omega_{i2}\} = \overline{A_{i1}}$, for instance. The elements of set system \mathcal{A}_i can be interpreted as *goal structures* as follows:

- A_{i1} - process goal of process component A_i has been achieved,
- $\overline{A_{i1}}$ - process goal of process component A_i has not been achieved,
- Ω_i - any goal of process component A_i has been achieved,
- \emptyset - nothing has been achieved.

The set \mathcal{A}_i is in sense of measure theory a σ -algebra. A σ -algebra is an algebraic structure which is closed against set-theoretical operations with its elements like union, intersection and complement set. That means, set-theoretical or logical operations with the elements of a σ -algebra will not cause any logical confusions.

We denote the elements of \mathcal{A}_i as *goal structures* and the set \mathcal{A}_i itself as *goal algebra* of process component A_i . In this context we will say: the goal in sense of a goal structure $A \in \mathcal{A}_i$ *has been achieved (is achieved)* if an $\omega \in A$ has been observed (is observed). The structure of a goal algebra \mathcal{A}_i is simple and has more a formal meaning here. The goal algebras are needed if we will go over to the corresponding product space for a common description of process components A_1, \dots, A_k .

3. Let $Q_i : \mathcal{A}_i \rightarrow [0, 1]$ be an additive function from our goal algebra \mathcal{A}_i into the interval $[0, 1]$ where to any given real number q_i , $0 \leq q_i \leq 1$,

$$Q_i(A_{i1}) = q_i, \quad Q_i(\overline{A_{i1}}) = 1 - q_i \quad \text{and} \quad Q_i(\Omega_i) = 1$$

holds. The values of numbers q_i and $1 - q_i$ are or can be interpreted as evaluation values for the goal structures A_{i1} and \bar{A}_{i1} . In this sense the values q_i and $1 - q_i$ form an evaluation distribution over the goal algebra \mathcal{A}_i . The function Q_i is a normalised measure on $(\Omega_i, \mathcal{A}_i)$. We denote $Q_i(A_{i1})$ as *score* that the goal of process component A_i has been achieved (is achieved), analogously $Q_i(\bar{A}_{i1})$ is the corresponding score that the goal in sense of process component A_i has not been achieved (is not achieved).

II. The product space. The triples $(\Omega_i, \mathcal{A}_i, Q_i)$, $i = 1, \dots, k$, are elementary measure spaces. We now will combine the so obtained spaces for the process components A_i to a product space. By this product space an evaluation of more complex goal structures becomes possible which include the goals of process components A_1, \dots, A_k . For the measure theoretical background used here we refer to [6], for instance.

The product space consists again of three objects Ω, \mathcal{A} and Q which are defined now as follows.

1. Let $\Omega = \Omega_1 \times \dots \times \Omega_k$ be the cross product over the goal spaces $\Omega_1, \dots, \Omega_k$. The elements of Ω are $\omega = \{\omega_1, \dots, \omega_k\}$ with $\omega_i \in \Omega_i$ for $i = 1, \dots, k$. We denote these elements as *k-dimensional elementary goals* and Ω is then the *k-dimensional goal space*.
2. Let \mathcal{A} be the set of all subsets of *k-dimensional goal space* Ω . This set of subsets forms again a σ -algebra over the goal space Ω . The elements of σ -algebra \mathcal{A} are denoted as *k-dimensional goal structures*. The σ -algebra \mathcal{A} is then the *k-dimensional goal algebra*.

Some examples of goal structures:

$$\begin{aligned}
 A &= \{\{\omega_{11}, \omega_{21}, \dots, \omega_{k1}\}\} - \text{all single goals } A_{11}, \dots, A_{k1} \text{ in} \\
 &\quad \text{sense of processes } A_1, \dots, A_k \text{ have been achieved (are} \\
 &\quad \text{achieved),} \\
 B &= \{\{\omega_{11}, \dots, \omega_{k-11}, \omega_{k1}\}, \{\omega_{11}, \dots, \omega_{k-11}, \omega_{k2}\}\} - \text{the} \\
 &\quad \text{single goals } A_{11}, \dots, A_{k-11} \text{ have been achieved (are} \\
 &\quad \text{achieved),} \\
 C &= \{\{\omega_{11}, \omega_{21}, \dots, \omega_{k-11}, \omega_{k2}\}\} - \text{the goals } A_{11}, \dots, \\
 &\quad A_{k-11} \text{ have been achieved, but not the goal } A_{k1},
 \end{aligned}$$

$$\begin{aligned}
D &= \{\{\omega_{11}, \omega_{22}, \dots, \omega_{k2}\}\} - \text{only the single goal } A_{11} \text{ has} \\
&\quad \text{been achieved (is achieved),} \\
E &= \{\{\omega_{11}, \omega_{2i_2}, \dots, \omega_{ki_k}\}, i_2, \dots, i_k \in \{1, 2\}\} - \text{the single} \\
&\quad \text{goal } A_{11} \in \mathcal{A}_i \text{ has been achieved.} \\
&\quad \text{It holds } E = A_{11} \times \Omega_2 \times \dots \times \Omega_k, A_{11} = \{\omega_{11}\} \in \mathcal{A}_1.
\end{aligned}$$

The last goal structure E is a special goal structure which is directed to the goal of process A_1 . Our goal algebra \mathcal{A} contains of k such goal structures which are given by

$$\begin{aligned}
A_1 &= A_{11} \times \Omega_2 \times \dots \times \Omega_k, & A_{11} &= \{\omega_{11}\} \in \mathcal{A}_1, & (6) \\
A_2 &= \Omega_1 \times A_{21} \times \dots \times \Omega_k, & A_{21} &= \{\omega_{21}\} \in \mathcal{A}_2, \\
&\dots \\
A_k &= \Omega_1 \times \Omega_2 \times \dots \times A_{k1}, & A_{k1} &= \{\omega_{k1}\} \in \mathcal{A}_k.
\end{aligned}$$

We denote these goal structures as *single goal structures*.

Beside of the single goal structures two further classes of goal structures are of special interest. We begin with the class of parallel goal structures. A goal structure B is said to be a *parallel goal structure* if a subset of single goal structures $A_1, \dots, A_s \in \mathcal{A}$ exists such that

$$B = \bigcup_{j=1}^s A_j$$

holds. A parallel goal structure B achieves its goal if at least one of the included single goal structures A_1, \dots, A_s reaches its goal. Our goal algebra \mathcal{A} contains $2^k - 1$ different parallel goal structures. These are

$$\begin{aligned}
&A_1, \dots, A_k, \quad A_1 \cup A_2, \dots, A_{k-1} \cup A_k, \quad A_1 \cup A_2 \cup A_3, \dots \\
&\dots, \quad \bigcup_{i=1}^k A_i = \Omega.
\end{aligned}$$

Because of the set system \mathcal{A} as the set of all subsets of Ω is a σ -algebra it holds:

$$\begin{aligned}
\text{(i)} \quad A_1, \dots, A_r \in \mathcal{A} &\Rightarrow \bigcup_{i=1}^r A_i \in \mathcal{A}, \\
\text{(ii)} \quad A \in \mathcal{A} &\Rightarrow \overline{A} \in \mathcal{A}.
\end{aligned}$$

This implies moreover

$$A_1, \dots, A_r \in \mathcal{A} \Rightarrow \bigcap_{i=1}^r A_i \in \mathcal{A}.$$

So beside the union of goal structures and the complement of a goal structure also the intersection of goal structures is again a goal structure. That means, the goal algebra \mathcal{A} is logical closed or consistent with respect to application of set theoretical operations to goal structures.

Beside parallel goal structures the following goal structures are of special interest. Let $B \in \mathcal{A}$ be a goal structure which to given single goal structures $A_1, \dots, A_r \in \mathcal{A}$ is defined by

$$B = \bigcap_{i=1}^r A_i.$$

Then the goal structure B achieves its goal if and only if when all single goal structures A_1, \dots, A_r achieve their goal. In analogy to definition of a parallel goal structure we denote the goal structure B as a *series goal structure*.

Finally, depending on to which process components two goal structure $B_1, B_2 \in \mathcal{A}$ refer we will say the goal structures B_1 and B_2 are said to be *separate goal structures* if the associated sets of process components are disjoint.

3. The product measure. Let $Q : \mathcal{A} \rightarrow [0, 1]$ be a map from the goal algebra \mathcal{A} into the interval $[0, 1]$ with the following property. For any goal structure $A = A_1 \times \dots \times A_k \in \mathcal{A}$ with $A_i \in \mathcal{A}_i, i = 1, \dots, k$, it holds

$$Q(A) = \prod_{i=1}^k Q_i(A_i). \quad (7)$$

Then, in sense of measure theory, Q is the product measure of measures $Q_i, i = 1, \dots, k$. This is according to the Hahn-Kolmogorov-Theorem of general measure theory a unique defined measure on measurable space (Ω, \mathcal{A}) .

Hence by the product measure Q a measure value $Q(A)$ is defined for each goal structure $A \in \mathcal{A}$. The value $Q(A)$ can be interpreted then as an evaluation number for it how the goal in sense of goal structure A has been achieved (can be achieved). In this sense big values $Q(A) \approx 1$ are a hint that the goal in sense of goal structure A has been achieved essentially, whereas a value $Q(A) \approx 0$ is a signal that goal in sense of goal structure A has been failed essentially. In this sense we will denote $Q(A)$ as *score for it that the goal of goal structure A has been achieved (can be achieved)* or, more short, simply as the *score of goal structure A* .

Collecting this together, the triple (Ω, \mathcal{A}, Q) forms a corresponding product space which allows an evaluation of all goal structures of goal algebra \mathcal{A} by means of the score measure Q defined by (7).

3.2 Calculation rules for scores

We will consider now calculation rules for the computation of scores of goal structures. The score Q defined by (7) is a normalised measure on (Ω, \mathcal{A}) . Each normalised measure possesses the following basic properties.

1. **Additivity.** According to the addition axiom of measure theory the following rule holds. Let $A_1, \dots, A_n \in \mathcal{A}$ be pairwise disjoint goal structures such that $A_i \cap A_j = \phi$ for $i \neq j$ and $i, j = 1, \dots, n$ holds. Then we have

$$Q\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n Q(A_i).$$

2. **Normalisation rule.** It holds

$$Q(\Omega) = 1.$$

From these rules further calculation rules result. The most important rules are the following. These are standard properties of each normalised measure. They hold correspondingly for the score measure Q considered here. We give a short survey.

3. **Complement rule.** Let $A \in \mathcal{A}$ be an arbitrary goal structure. Then it holds

$$Q(A) = 1 - Q(\bar{A}).$$

Proof. This is a standard property of any normalised measure. \square

4. General addition rule. Let $A_1, \dots, A_r \in \mathcal{A}$ be arbitrary goal structures. Then it holds

$$Q(A_1 \cup A_2) = Q(A_1) + Q(A_2) - Q(A_1 \cap A_2)$$

as well as

$$\begin{aligned} Q\left(\bigcup_{i=1}^r A_i\right) &= \sum_{i=1}^r Q(A_i) - \sum_{\substack{i,j=1 \\ i < j}}^r Q(A_i \cap A_j) + \sum_{\substack{i,j,k=1 \\ i < j < k}}^r Q(A_i \cap A_j \cap A_k) \\ &\quad - \dots + (-1)^{r+1} Q(A_1 \cap \dots \cap A_r). \end{aligned}$$

Proof. This rule corresponds the general addition rule of measure theory. \square

Now we will consider some calculation rules which are of particular interest for evaluation of special goal structures.

5. Product rule for series goal structures (series rule). Let $A_1, \dots, A_r \in \mathcal{A}$ be single goal structures in sense of relation (6) with $Q(A_i) = q_i$, $0 \leq q_i \leq 1$, $i = 1, \dots, r$, $r \leq k$. Let $B = \bigcap_{i=1}^r A_i$ be a corresponding series goal structure. Then we have

$$Q(B) = Q\left(\bigcap_{i=1}^r A_i\right) = \prod_{i=1}^r Q(A_i) = \prod_{i=1}^r q_i.$$

Proof. Without of any loss of generality we suppose that the single goal structures $A_1, \dots, A_r \in \mathcal{A}$ are directed to the first r processes components A_1, \dots, A_r . Then we have

$$\bigcap_{i=1}^r A_i = A_{11} \times \dots \times A_{r1} \times \Omega_{r+1} \times \dots \times \Omega_k \quad \text{with} \quad A_{i1} \in \mathcal{A}_i, \quad i = 1, \dots, r.$$

By (7) we get

$$Q\left(\bigcap_{i=1}^r A_i\right) = Q_1(A_{11}) \cdots Q_r(A_{r1}) Q_{r+1}(\Omega_{r+1}) \cdots Q_k(\Omega_k). \quad (8)$$

For single goal structures A_i holds $Q(A_i) = Q_i(A_{i1})$. With $Q(A_i) = q_i$ for $i = 1, \dots, r$ and $Q_i(\Omega_i) = 1$ for $i = r + 1, \dots, k$ this implies

$$Q\left(\bigcap_{i=1}^r A_i\right) = \prod_{i=1}^r Q_i(A_{i1}) = \prod_{i=1}^r Q(A_i) = \prod_{i=1}^r q_i. \quad \square$$

6. Addition rule for parallel goal structures (parallel rule). Let $A_1, \dots, A_r \in \mathcal{A}$ be single goal structures with $Q(A_i) = q_i$, $i = 1, \dots, r$, $r \leq k$. Let $B = \cup_{i=1}^r A_i$ be a corresponding parallel goal structure. Then it holds

$$Q(B) = Q\left(\bigcup_{i=1}^r A_i\right) = 1 - \prod_{i=1}^r (1 - q_i). \quad (9)$$

Proof. By means of the complement rule and de Morgan's rule we obtain

$$Q\left(\bigcup_{i=1}^r A_i\right) = Q\left(\overline{\bigcap_{i=1}^r \overline{A_i}}\right) = 1 - Q\left(\bigcap_{i=1}^r \overline{A_i}\right) = 1 - Q\left(\bigcap_{i=1}^r \overline{A_i}\right).$$

With product measure property of Q and again by the complement rule we get

$$Q\left(\bigcap_{i=1}^r \overline{A_i}\right) = \prod_{i=1}^r Q(\overline{A_i}) = \prod_{i=1}^r (1 - Q(A_i)) = \prod_{i=1}^r (1 - q_i).$$

Collecting this together we get relation (9). □

A special case is the case $r = 2$. Then we have

$$Q(A_1 \cup A_2) = 1 - (1 - q_1)(1 - q_2) = q_1 + q_2 - q_1 q_2.$$

For $r = 3$ rule (9) modifies to

$$Q(A_1 \cup A_2 \cup A_3) = q_1 + q_2 + q_3 - q_1 q_2 - q_1 q_3 - q_2 q_3 + q_1 q_2 q_3.$$

The addition rule for parallel goal structures is a special case of general addition rule. This is the reason why we denote this rule as an addition rule too.

7. Product rule for generalised series goal structures (generalised series rule). Let $B_1, \dots, B_r \in \mathcal{A}$ be r , $1 \leq r \leq k$, separate parallel goal structures with

$$B_i = \bigcup_{j=1}^{s_i} A_{ij}, \quad (10)$$

generated by s_i single goal structures $A_{ij} \in \mathcal{A}$ with $Q(A_{ij}) = q_{ij}$ for $i = 1, \dots, r$, $j = 1, \dots, s_i$, $\sum_{i=1}^r s_i = k$. Let $C \in \mathcal{A}$ be a goal structure defined by

$$C = \bigcap_{i=1}^r B_i = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}. \quad (11)$$

Then it holds

$$\begin{aligned} Q(C) &= Q\left(\bigcap_{i=1}^r B_i\right) = Q\left(\bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}\right) = \prod_{i=1}^r Q\left(\bigcup_{j=1}^{s_i} A_{ij}\right) \\ &= \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij})\right). \end{aligned} \quad (12)$$

Proof. For each parallel goal structure $B_i = \cup_{j=1}^{s_i} A_{ij}$, $i = 1, \dots, r$, we consider the corresponding product measure space $(\Omega^{(i)}, \mathcal{A}^{(i)}, Q^{(i)})$ generated by the measure spaces $(\Omega_{ij}, \mathcal{A}_{ij}, Q_{ij})$, $j = 1, \dots, s_i$. As above, the product measures $Q^{(i)}$ are defined by

$$Q^{(i)}(A_{i1} \times \dots \times A_{is_i}) = Q_{i1}(A_{i1}) \cdots Q_{is_i}(A_{is_i}), \quad A_{ij} \in \mathcal{A}_{ij}, \quad j = 1, \dots, s_i.$$

Let $A_j^{(i)}$ be a single goal structures of goal algebra $\mathcal{A}^{(i)}$ which is directed to the goal of process component A_{ij} with $Q^{(i)}(A_j^{(i)}) = q_{ij}$, $j = 1, \dots, s_i$. Then, according the addition rule for a parallel goal structure $B_i = \cup_{j=1}^{s_i} A_j^{(i)} \in \mathcal{A}^{(i)}$ we get

$$Q^{(i)}(B_i) = Q^{(i)}\left(\bigcup_{j=1}^{s_i} A_j^{(i)}\right) = 1 - \prod_{j=1}^{s_i} (1 - q_{ij}). \quad (13)$$

If we now consider the product of measure spaces $(\Omega^{(1)}, \mathcal{A}^{(1)}, Q^{(1)}), \dots, (\Omega^{(r)}, \mathcal{A}^{(r)}, Q^{(r)})$ then we obtain again the product measure space (Ω, \mathcal{A}, Q) which is generated by the measure spaces $(\Omega_{ij}, \mathcal{A}_{ij}, Q_{ij})$ with $i = 1, \dots, r$ and $j = 1, \dots, s_i$.

The corresponding product measure Q is then defined as follows. For any goal structure $C = B_1 \times \dots \times B_r \in \mathcal{A}$, $B_i \in \mathcal{A}^{(i)}$, $i = 1, \dots, r$, we have

$$Q(C) = Q(B_1 \times \dots \times B_r) = Q^{(1)}(B_1) \dots Q^{(r)}(B_r).$$

This, together with (13) completes the proof. \square

8. Product rule for complex series processes (consistency rule). Let A'_{ij} be for $i = 1, \dots, r$ and $j = 1, \dots, s_i$ separate processes to a given set of process components \mathcal{G} . Let C' be a series process whose elements are separate parallel processes B'_i which consist of separate processes $A'_{i1}, \dots, A'_{is_i}$. Denote by C', B'_1, \dots, B'_r and $A'_{11}, \dots, A'_{rs_r}$ the corresponding goal structures. Then it holds

$$C' = \bigcap_{i=1}^r B'_i = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A'_{ij}$$

and the score that process C' achieves its goal is

$$Q(C') = \prod_{i=1}^r Q(B'_i) = \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - Q(A'_{ij})) \right). \quad (14)$$

Proof. This is in case of separateness of included processes A'_{ij} again a property of the product measure Q . \square

This rule is an extension of generalised series rule which allows to include processes instead of process components at evaluation of process goals. By this a consistent evaluation of complex process goal structures becomes possible. Hence rule (14) can be considered as main rule for process evaluation. By this rule the score calculation of a goal structure C can be reduced consistent to scores for goal structures of separate processes.

The side condition of separateness of involved processes is in many cases fulfilled by the application background. If there are intersections between the involved processes then it is possible to switch over to separate processes

by consideration of finer process structures. This increases a bit the calculation effort. But it is a finite procedure because we have a finite number of process components.

3.3 Estimation of scores for goal structures

In the previous section it has been shown how scores can be calculated for complex goal structures. By the calculation rules considered there the scores can be reduced to the scores of the single goal structures of goal algebra \mathcal{A} . Unfortunately, as a rule these scores are not given a-priori and we need corresponding estimation methods for these scores as well as for the scores of other goal structures of \mathcal{A} .

We assume that the considered elementary processes A_1, \dots, A_k can be observed via ordinal or metrical ordered observation variables X_1, \dots, X_k . Let $\mathcal{X}_i = [x'_i, x''_i]$ be the domain or scale of i -th observation variable X_i , $i = 1, \dots, k$. High values of X_i in the neighbourhood of upper bound x''_i of evaluation scale are an indication of that the goal of process A_i has been achieved, essentially. Small values in the neighbourhood of x'_i a corresponding signal that the process goal has been failed, essentially. Observation values $X_i = x''_i$ or $X_i = x'_i$ indicate that the goal of process A_i has been completely achieved or failed, respectively. The scales which are used for observation variables X_i can be continuous or discrete, must be ordered and can be, for instance, rank places too.

For explanation of principal approach at estimation of score of a goal structure we consider a generalised series goal structure

$$C = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}.$$

The score $Q(C)$ of this goal structure is calculated according to the generalised series rule (12) by

$$Q(C) = \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}) \right).$$

Let $A_{11}, \dots, A_{r s_r}$ be the elementary processes which contribute to achievement of goal of goal structure C , let $\vec{X} = (X_{11}, \dots, X_{r s_r})$ be the vector of

observation variables by which the processes A_{11}, \dots, A_{rs_r} can be observed and let $[x'_{ij}, x''_{ij}]$, $x'_{ij} < x''_{ij}$, be the observation interval of variable X_{ij} .

Now we assume that we have a sample of size n , $n \geq 1$, to our observation vector \vec{X} . Such a sample can be obtained, e.g., by a corresponding interrogation of participants of an e-learning program via an assessment checklist after the course is finished. This would be an a-posteriori-interrogation. Or, one could interrogate experts which evaluate the course based on the course materials before the course is held. This would be an a-priori-interrogation.

Let $\vec{x}^{(k)} = (x_{11}^{(k)}, \dots, x_{rs_r}^{(k)})$ be the k^{th} sampling element in a sample of size n , $k = 1, \dots, n$. Then at first we have to normalise our sample values $x_{ij}^{(k)}$ by transforming of these values to the interval $[0, 1]$, the domain for our scores q_{11}, \dots, q_{rs_r} . This is reached by the following transformation. Let $q_{ij}^{*(k)}$ for $k = 1, \dots, n$, $i = 1, \dots, r$ and $j = 1, \dots, s_i$ be defined by

$$q_{ij}^{*(k)} = \frac{x_{ij}^{(k)} - x'_{ij}}{x''_{ij} - x'_{ij}}. \quad (15)$$

Then $q_{ij}^{*(k)}$ is an estimation value for score q_{ij} based on observation value $x_{ij}^{(k)}$. It holds $0 \leq q_{ij}^{*(k)} \leq 1$.

With the so obtained estimation values $q_{ij}^{*(k)}$ for q_{ij} we get an estimation value for $Q(C)$ if we substitute in formula (12) of Section 3.2 the unknown scores q_{ij} by the estimation values $q_{ij}^{*(k)}$. We get

$$Q^{*(k)}(C) = \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)}) \right) \quad (16)$$

for $k = 1, \dots, n$.

Hence, collecting these values together we obtain a sample $(Q^{*(1)}(C), \dots, Q^{*(n)}(C))$ of size n for score $Q(C)$. By the method of moments we obtain finally via the arithmetic mean an estimation function for the score $Q(C)$ by

$$Q^*(C) = \frac{1}{n} \sum_{k=1}^n Q^{*(k)}(C) = \frac{1}{n} \sum_{k=1}^n \prod_{i=1}^r \left(1 - \prod_{j=1}^{s_j} (1 - q_{ij}^{*(k)}) \right). \quad (17)$$

This is the main formula for an estimation of score of a generalised series goal structure based on a sample of size n in context of an interrogation.

Special cases are again the parallel and series goal structures. For a parallel goal structure $B_i = \cup_{j=1}^{s_i} A_{ij}$ we get according the addition rule for parallel goal structures as estimation function for $Q(B_i)$:

$$Q^*(B_i) = Q^*\left(\bigcup_{j=1}^{s_i} A_{ij}\right) = \frac{1}{n} \sum_{k=1}^n \left(1 - \prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)})\right). \quad (18)$$

For series goal structures $C_i = \cap_{j=1}^r A_{ij}$ we get by the product rule for series goal structures as estimation for score $Q(C_i)$:

$$Q^*(C_i) = Q^*\left(\bigcap_{i=1}^r A_{ij}\right) = \frac{1}{n} \sum_{k=1}^n \prod_{j=1}^r q_{ij}^{*(k)}. \quad (19)$$

The score $q_{ij} = Q(A_{ij})$ of a single goal structure $A_{ij} \in \mathcal{A}$ can be estimated by

$$Q^*(A_{ij}) = \frac{1}{n} \sum_{k=1}^n q_{ij}^{*(k)}. \quad (20)$$

In case of missing values in the sample the missing values for $q_{ij}^{*(k)}$ can be substituted then by the estimation values $Q^*(A_{ij})$ which are obtained via the arithmetic mean based on the incomplete sample. This corresponds a 'neutral' evaluation of missing values in the sample.

3.4 Calibration rules for empirical scores

Depending on complexity of a logical goal structure the associated score value can be quite small. That can happen even in situations where the scores for involved process components are high. That is a consequence of product measure which forms the base for evaluation of a multidimensional goal structure.

If we have, for instance, a series goal structure which consists of five key goals B_i with $Q(B_i) = q = 0.8$ for $i = 1, \dots, 5$ then we get for the score $Q(C)$ that the aim of goal structure $C = \cap_{i=1}^5 B_i$ is achieved the value

$$Q(C) = Q(\cap_{i=1}^5 B_i) = 0.8^5 = 0.3277.$$

Whether the key goal scores $Q(B_i)$ are quite high the obtained score $Q(C)$ is surprisingly small. If all key goals have an average score $Q(B_i) = 0.5$ for $i = 1, \dots, 5$ we get even $Q(C) = 0.5^5 = 0.03125$. That means, the 'average level' of a goal structure $C = \bigcap_{i=1}^5 B_i$ is scored by a comparatively small score value 0.0313. Although this is a general property of any product measure, which follows from the corresponding product rule, such small values can cause irritations especially when the evaluators are not so familiar with measure theoretical background of SURE model.

Hence a further transformation of obtained score values would be desirable by which the following is reached:

1. The obtained score possesses a 'normal' value which allows a further interpretation of this value.
2. It becomes possibly to judge by the score value whether the aims of a goal structure have been reached in a quality above or below the average.
3. It should be possibly to compare score values which belong to different goal structures.

That can be reached by the following transformation or calibration of score value.

Definition. Let $C \in \mathcal{A}$ be a goal structure defined by simple goals $A_{ij} \in \mathcal{A}$, $i = 1, \dots, r, j = 1, \dots, s_i$, such that

$$C = \bigcap_{i=1}^r B_i = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij} \quad (21)$$

holds. Let $Q(A_{ij})$ be the score of A_{ij} . Then

$$Q_e(C) = \sqrt[r]{\prod_{i=1}^r \left(1 - \sqrt[s_i]{\prod_{j=1}^{s_i} (1 - Q(A_{ij}))} \right)} \quad (22)$$

is called *evaluation score (calibrated score)* of goal structure C .

The evaluation score is a generalised geometric mean. By means the evaluation score a standard score value $Q(C)$ is transformed into a better interpretable score value $Q_e(C)$.

The evaluation score $Q_e(C)$ possesses the following properties.

1. Average value rule. If $Q(A_{ij}) = 0.5$ holds for $i = 1, \dots, r$ and $j = 1, \dots, s_i$, that means all single goals A_{ij} are achieved with average quality, then we have

$$Q_e(C) = 0.5.$$

This immediately follows from the definitions of $Q(C)$ and $Q_e(C)$.

If all single goals are achieved with average quality then a logical goal structure can reach its goal also only in average quality. Hence a value $Q_e(C) = 0.5$ is an indication for evaluator that a goal structure C has achieved its aim in average quality. It is interesting in this context that this average value 0.5 doesn't depend on inner structure of a logical goal structure. Hence the value 0.5 becomes to a reference value.

By means of that reference value 0.5 we have a possibility to recognise whether a goal structure has reached its aims better or worse than average. By this it becomes possible to compare different goal structures in this sense too.

Moreover, if $Q(A_{ij}) = q$ holds for any $0 \leq q \leq 1$, $i = 1, \dots, r$ and $j = 1, \dots, s_i$, then we get for a goal structure $C = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}$ according relation (21)

$$Q_e(C) = q.$$

This property can be interpreted as follows. An evaluation score value $Q_e(C) = q$, $0 \leq q \leq 1$, mirrors an 'average satisfaction' with the result in sense of goal structure C at the rate of $q \cdot 100\%$. For instance, if we have an e-learning course, which is evaluated by students in view of a given goal structure with an evaluation value $Q_e(C) = 0.85$ then we can regard this as follows. The average satisfaction level of students with course amounts to 85%.

2. Zero and one continuity. If at least one key goal B_i , $i = 1, \dots, r$ doesn't reach its aim, that means it exists at least one index $i^* \in \{1, \dots, r\}$ such that

$Q(B_{i^*}) = 0$ holds, then we have

$$Q_e(C) = Q_e(\bigcap_{i=1}^r B_i) = 0.$$

If all scores $Q(B_i) = 1$ for $i = 1, \dots, r$ then we have

$$Q_e(C) = Q_e(\bigcap_{i=1}^r B_i) = 1.$$

3. Monotonicity of evaluation score function. Let $C = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}$ be a given goal structure. Let $(q_{11}, \dots, q_{rs_r})$ and $(q'_{11}, \dots, q'_{rs_r})$ be two score value sets for $(A_{11}, \dots, A_{rs_r})$ with

$$q_{ij} \leq q'_{ij} \quad \text{for } i = 1, \dots, r \quad \text{and } j = 1, \dots, s_r$$

and let $Q_e(C)$ and $Q'_e(C)$ be the corresponding evaluation scores. Then we have

$$Q_e(C) \leq Q'_e(C).$$

The properties 2 and 3 are direct consequences of the corresponding rules as they hold for the non-calibrated score function Q .

4. The conservativity of evaluation score function. The evaluation score function is conservative in the following sense.

a) Let $C = B_1 \cap B_2$ be a logical series structure with $Q(B_1) = 0.5 - \varepsilon$ and $Q(B_2) = 0.5 + \varepsilon$ for any $\varepsilon, 0 < \varepsilon < 0.5$. Then we have

$$\frac{1}{2}(Q(B_1) + Q(B_2)) = 0.5.$$

That means, the 'average' quality that the goals B_1 and B_2 are achieved is 0.5. Good is in this context of course the value $Q(B_2) = 0.5 + \varepsilon > 0.5$, not so good the value $Q(B_1) = 0.5 - \varepsilon < 0.5$. That means, single small values for $Q(B_i)$ can become critical in context with evaluation of whole goal structure.

The evaluation score function Q_e behaves itself conservative in such situations. We get

$$Q_e(C) = \sqrt{Q(B_1)Q(B_2)} = \sqrt{(0.5 - \varepsilon)(0.5 + \varepsilon)} = \sqrt{0.5^2 - \varepsilon^2} < 0.5.$$

Hence, the goal structure $C = B_1 \cap B_2$ is evaluated as below-average which is a right reaction is in this case.

b) Let $C = B_1 \cup B_2$ be a logical parallel structure with $Q(B_1) = 0.5 - \varepsilon$ and $Q(B_2) = 0.5 + \varepsilon$ again. Then in sense of arithmetic mean the 'average' quality by which the goals B_1 and B_2 are reached is again 0.5. For a parallel structure the more welcome value is now the value $Q(B_2) = 0.5 + \varepsilon > 0.5$. That means the parallel structure will reach its aim in an 'above-average' quality. This is mapped correctly by the evaluation score function. Now we have

$$\begin{aligned} Q_e(C) &= Q_e(B_1 \cup B_2) = 1 - \sqrt{(1 - Q(B_1))(1 - Q(B_2))} \\ &= 1 - \sqrt{(1 - (0.5 - \varepsilon))(1 - (0.5 + \varepsilon))} \\ &= 1 - \sqrt{0.5^2 - \varepsilon^2} > 0.5. \end{aligned}$$

That confirms that the evaluation score function reacts again in the right direction.

Collecting this together the evaluation score function Q_e reacts correct as well at logical series as at logical parallel goal structures with respect to deviations of included scores from average value 0.5. This behaviour of evaluation score function Q_e we denote as 'conservative' behaviour of Q_e against deviations of included scores from the average value 0.5. This conservative behaviour holds analogously if we consider deviations from a score value $q_0 \neq 0.5$ with $Q(B_1) = q_0 - \varepsilon$ and $Q(B_2) = q_0 + \varepsilon$.

5. Estimation of evaluation scores. In analogous manner like at estimation of score of a goal structure based on observation results of an adapted checklist the evaluation score can be estimated as follows.

We suppose we have a sample of size n of corresponding checklist results. Let according formula (15) of Section 3.3 $q_{ij}^{*(k)}$ be for $i = 1, \dots, r, j = 1, \dots, s_i$ and $k = 1, \dots, n$ the estimation value for $Q(A_{ij})$ based on the k^{th} checklist data record. Then we obtain for each checklist record an estimation for $Q_e(C)$ with

$$C = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}$$

if we substitute in formula (22) the unknown score values $Q(A_{ij})$ by the estimation values $q_{ij}^{*(k)}$. For each of the n data records of our checklist

sample we get an estimation value for $Q_e(C)$ by

$$Q_e^{*(k)}(C) = \sqrt[r]{\prod_{i=1}^r \left(1 - s_i \sqrt[s_i]{\prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)})} \right)},$$

$k = 1, \dots, n$. By the method of moments we get then via the arithmetic mean over the values $Q_e^{*(1)}(C), \dots, Q_e^{*(n)}(C)$ an estimation value for evaluation score $Q_e(C)$.

Definition. To a given goal structure

$$C = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}$$

and a sample of size n of checklist results the value

$$Q_e^*(C) = \frac{1}{n} \sum_{k=1}^n Q_e^{*(k)}(C) = \frac{1}{n} \sum_{k=1}^n \sqrt[r]{\prod_{i=1}^r \left(1 - s_i \sqrt[s_i]{\prod_{j=1}^{s_i} (1 - q_{ij}^{*(k)})} \right)} \quad (23)$$

is denoted as *empirical evaluation score* for goal structure C .

This is the main formula for estimation of evaluation scores of logical oriented goal structures.

6. Confidence intervals for evaluation scores. Checklist results depend always on random influences which work in the background. Hence a checklist result like $(Q_e^{*(1)}(C), \dots, Q_e^{*(n)}(C))$ can be considered as a sample of size n for the evaluation score $Q_e(C)$. In this sense the evaluation score $Q_e(C)$ is a random variable. Then like any random variable the evaluation score $Q_e(C)$ has an expectation value as well as a variance. Denote by

$$\tilde{q} = EQ_e(C) \quad \text{and} \quad \tilde{\sigma}^2 = D^2 Q_e(C)$$

the expectation value and the variance of $Q_e(C)$, respectively. Depending on the value of variance $\tilde{\sigma}^2$ an estimation value $Q_e^*(C)$ for $Q_e(C)$ is more or less precisely. To evaluate the precision of an empirical evaluation score

value we can consider a confidence interval for the expectation value \tilde{q} of evaluation score $Q_e(C)$.

A confidence interval for \tilde{q} is a random interval which covers the unknown score value \tilde{q} with a pre-given probability, the confidence level $1 - \alpha$. Frequently used values for the probability α are values like $\alpha = 0.1, 0.05$ or 0.01 . The construction of a confidence interval requires the knowledge of probability distribution of observation variable $Q_e(C)$. This distribution is because of the complexity of random background in the present case not known. Therefore we will consider an asymptotic confidence interval. This is a possibility nevertheless to get a confidence interval if the sample size is sufficiently large.

We proceed as follows. By the central limit theorem the empirical evaluation score

$$Q_e^*(C) = \frac{1}{n} \sum_{k=1}^n Q_e^{*(k)}(C)$$

is asymptotically, that means for $n \rightarrow \infty$, normal distributed with mean \tilde{q} and variance $\tilde{\sigma}^2/n$. This implies the standardised random variable

$$Z_n = \sqrt{n} \frac{Q_e^*(C) - \tilde{q}}{\tilde{\sigma}}$$

is asymptotically $N(0, 1)$ -distributed.

A confidence interval for \tilde{q} can be obtained then as follows. Let z_α^* be defined to a pre-given probability α , $0 < \alpha < 1$, by

$$\Phi(z_\alpha^*) = \int_{-\infty}^{z_\alpha^*} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 1 - \alpha/2.$$

Here $\Phi(z)$ denotes the distribution function and z_α^* the $(1 - \alpha/2)$ -quantile of standardised normal distribution. Then it holds with probability $1 - \alpha$ for sufficient large values of sample size n

$$-z_\alpha^* \leq Z_n \leq z_\alpha^*$$

or

$$Q_e^*(C) - z_\alpha^* \frac{\tilde{\sigma}}{\sqrt{n}} \leq \tilde{q} \leq Q_e^*(C) + z_\alpha^* \frac{\tilde{\sigma}}{\sqrt{n}}.$$

Definition. The interval

$$\left[Q_e^*(C) - z_\alpha^* \frac{\tilde{\sigma}}{\sqrt{n}}, Q_e^*(C) + z_\alpha^* \frac{\tilde{\sigma}}{\sqrt{n}} \right]$$

is said to be an *asymptotic two-sided confidence interval* for the average evaluation score value \tilde{q} at *confidence level* $1 - \alpha$.

This interval is to interpret statistically as follows. It covers in average in $(1 - \alpha) \cdot 100\%$ of all cases the unknown parameter value \tilde{q} .

The problem here is that the variance $D^2 Q_e(C) = \tilde{\sigma}^2$ is unknown. The usual approach in such cases is to substitute the unknown variance $\tilde{\sigma}^2$ by a corresponding estimation value for the unknown variance like

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (Q_e^{*(k)}(C) - Q_e^*(C))^2 \quad \text{with} \quad Q_e^*(C) = \frac{1}{n} \sum_{k=1}^n Q_e^{*(k)}(C).$$

The so obtained interval

$$\left[Q_e^*(C) - z_\alpha^* \frac{S}{\sqrt{n}}, Q_e^*(C) + z_\alpha^* \frac{S}{\sqrt{n}} \right]$$

is then an approximation for a two-sided asymptotic confidence interval for the evaluation score \tilde{q} at confidence level $1 - \alpha$ based on a sample of size n .

Beside a confidence interval the sample standard deviation

$$\tilde{\sigma}^* = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (Q_e^{*(k)}(C) - Q_e^*(C))^2}$$

is a statistical quantity describing the precision of a score estimation value.

3.5 Simulation examples

In this section we consider some simulation examples describing the behaviour of SURE model under several evaluation conditions. We assume

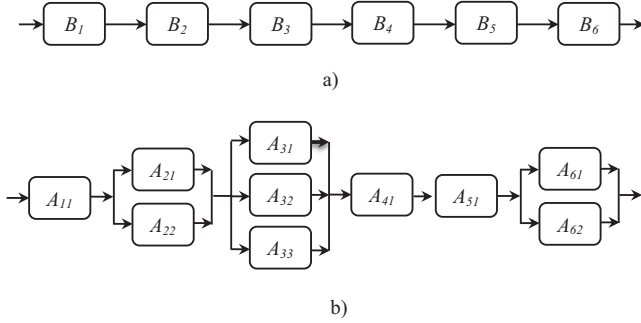


Figure 10: Goal structures for simulation examples.

that we have a goal structure

$$C = \bigcap_{i=1}^r B_i = \bigcap_{i=1}^r \bigcup_{j=1}^{s_i} A_{ij}$$

consisting of $r = 6$ key goals B_1, \dots, B_6 with

$$B_1 = A_{11}, \quad B_2 = A_{21} \cap A_{22}, \quad B_3 = A_{31} \cap A_{32} \cap A_{33}$$

and

$$B_4 = A_{41}, \quad B_5 = A_{51}, \quad B_6 = A_{61} \cap A_{62}.$$

The key goal structures B_1, B_4 and B_5 are single goal structures, the key goal structures B_2, B_3 and B_6 are parallel structures. The logical scheme of this goal structure is shown by Figure 10.

We assume that based on an adapted checklist a sample of size $n = 25$ has been obtained. The evaluation intervals are for all questions equal. The evaluation scale includes the numbers $0, 1, \dots, 10$. Zero means the goal has been failed, the value 10 stands for the goal has been achieved.

Example 1. The satisfied students.

We assume that students have been satisfied with course in that sense, that all questions of checklist have been answered with scores between 4 and 10. Table 4 contains the corresponding checklist results. The scores are realizations of a uniformly distributed random variable with values in the set $\{4, 5, \dots, 10\}$. In the centre of table we see the sample scores. They are highlighted according to a grey scale which is ranging from darkgrey over middle grey to white. White means goal has been achieved, middle grey corresponds an average evaluation, dark grey means goal was not achieved. Each row from 1 until 25 contains a checklist data record. The dominating grey values are white until middle grey, a first signal that the course is running well.

The third row from below contains the empirical scores $Q^*(A_{ij})$ of single goals A_{ij} . These scores are identical with the empirical evaluation scores $Q_e^*(A_{ij})$. Next line shows the empirical key goal evaluation scores $Q_e^*(B_i)$. These values are comparatively high. That confirms the satisfaction of students in view of achievement of a single key goal. The last row contains the empirical standard scores $Q^*(B_i)$ for key goals B_i .

In the second last and last column of table the empirical evaluation scores $Q_e^{*(k)}(C)$ and empirical standard scores $Q^{*(k)}(C)$ of checklist records for $k = 1, \dots, 25$ are displayed. The arithmetic means over these scores provide the empirical evaluation score $Q_e^*(C)$ and the empirical standard score $Q^*(C)$. The bold gray scaled value in the right lower corner is the empirical evaluation score $Q_e^*(C)$, on the right hand of that we have the empirical standard score $Q^*(C)$.

We get $Q_e^*(C) = 0.73$. This value is greater than 0.5. Hence the course is evaluated above average. The value $Q_e^*(C) = 0.73$ stands for a quite high satisfaction of students with course. It corresponds an average satisfaction level of 73%. The empirical score $Q^*(C)$ is comparatively small in comparison with evaluation score $Q_e^*(C)$. This is a peculiar characteristic of product measure.

The lower table of Table 4 presents confidence intervals for evaluation score at the confidence levels 0.90, 0.95 and 0.99. The confidence intervals are quite wide. That is a consequence of the not so high sample size of $n = 25$ here. No confidence interval contains the value 0.5. This is a hint

k	B_1		B_2		B_3			B_4	B_5	B_6		$Q_e^{*(k)}$	$Q^{*(k)}$
	A_{11}	A_{21}	A_{22}	A_{31}	A_{32}	A_{33}	A_{41}	A_{51}	A_{61}	A_{62}			
1	10	4	7	4	9	5	4	9	10	9	0.72	0.29	
2	7	6	7	4	7	5	8	6	6	4	0.63	0.20	
3	8	5	9	5	10	8	10	8	6	6	0.82	0.51	
4	8	9	6	6	10	5	4	6	9	7	0.71	0.18	
5	10	5	5	6	9	9	10	5	4	10	0.77	0.37	
6	5	9	10	7	10	7	9	6	4	6	0.70	0.21	
7	8	5	7	6	9	4	6	8	6	8	0.70	0.29	
8	6	7	8	4	8	8	10	8	5	7	0.74	0.37	
9	6	9	4	10	4	4	10	5	7	6	0.73	0.25	
10	4	6	8	6	5	7	10	5	7	7	0.63	0.16	
11	6	7	9	9	6	5	4	6	10	4	0.67	0.14	
12	6	5	7	4	8	7	5	10	9	6	0.68	0.24	
13	5	8	10	8	8	10	9	9	4	8	0.80	0.36	
14	7	8	6	7	4	8	10	9	6	6	0.75	0.47	
15	10	9	6	6	8	9	5	6	4	5	0.67	0.20	
16	7	9	8	6	8	6	7	9	5	8	0.75	0.38	
17	4	7	7	6	5	4	5	9	6	8	0.60	0.13	
18	7	9	8	10	5	9	7	9	7	8	0.81	0.41	
19	9	10	4	7	4	7	5	7	6	4	0.68	0.23	
20	7	9	10	8	6	5	10	10	5	5	0.78	0.50	
21	9	10	7	8	6	6	4	8	6	5	0.69	0.22	
22	9	5	7	4	8	9	6	4	8	10	0.68	0.18	
23	10	9	9	6	8	5	10	8	7	5	0.81	0.65	
24	5	4	10	8	7	9	10	5	8	9	0.75	0.24	
25	10	10	10	5	10	7	10	6	7	10	0.92	0.60	
$Q^*(A_{ij})$	0.73	0.74	0.76	0.64	0.73	0.67	0.75	0.72	0.65	0.68	0.73	0.31	
$Q_e^*(B_i)$	0.73	0.80		0.78			0.75	0.72	0.71				
$Q^*(B_i)$	0.73	0.94		0.97			0.75	0.72	0.89				

Scale										
0	1	2	3	4	5	6	7	8	9	10
0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

Confidence intervals				
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e
0.90		0.70	0.75	
0.95	0.73	0.70	0.76	0.07
0.99		0.69	0.77	

Table 4: Checklist results, scores and confidence intervals - Example 1.

that the deviation of obtained empirical evaluation score 0.73 from value 0.5 can be considered as significant.

In summary the sample result indicates a significant above-average satisfaction of students with course in an average rate of approximately 73%. The empirical evaluation scores for key goals have values near the empirical evaluation score for total goal. That means, the achievement of key goals is evaluated in the same direction like the achievement of total course goal. There is no evidence for particular weak points in view of the key goals to be reached.

Example 2. The unsatisfied students.

This is the counter example to Example 4. The students are unsatisfied in broad front with course. The score set is again the set $\{0, 1, \dots, 10\}$. The students evaluate the success of single goals between 0 and 6. Table 5 shows the checklist results. The scores are now realizations of a uniformly distributed random variable with values in the set $\{0, 1, \dots, 6\}$. The dominating grey values are dark grey until middle grey. This is a first hint that something went wrong with that course.

The values for the empirical evaluation scores of key goals B_1, \dots, B_6 are quite high yet. They range from 0.28 until 0.35, but the empirical total evaluation score $Q_e^*(C)$ is only 0.17. This is a comparatively small value and seems to be a contradiction to the values of empirical scores for key goals. In fact this value shows that the SURE model is working correct. The SURE model includes the whole logical structure of total goal into evaluation. This goal is failed if only one of key goals has been failed. The second column from right shows that the total goal has been failed in that sense in 11 from 25 cases. This is finally the reason for an empirical evaluation score $Q_e^*(C)$ of only 0.17.

This behaviour of SURE model is an essential advantage against linear evaluation model. Linear models react in such cases quite insensible. They are not able to recognise such situations as a rule.

Example 3. The gambling students.

This is a fictive situation where we assume that students do not answer to question of checklist but select randomly a score value from $\{0, 1, \dots, 10\}$ as answer.

k	B_1		B_2		B_3			B_4	B_5	B_6		$Q_e^{*(k)}$	$Q^{*(k)}$
	A_{11}	A_{21}	A_{22}	A_{31}	A_{32}	A_{33}	A_{41}	A_{51}	A_{61}	A_{62}			
1	6	1	1	5	1	4	1	6	1	1	0.22	0.00	
2	2	5	6	4	3	5	5	5	6	4	0.42	0.02	
3	5	2	1	5	2	2	4	0	3	4	0.00	0.00	
4	6	5	4	6	2	5	5	0	3	3	0.00	0.00	
5	2	0	4	4	2	4	4	4	2	5	0.31	0.01	
6	3	2	1	6	0	0	2	3	3	4	0.25	0.00	
7	6	5	3	2	0	0	2	1	0	6	0.22	0.00	
8	5	2	2	2	6	6	5	0	3	6	0.00	0.00	
9	5	0	6	2	3	6	4	4	2	6	0.41	0.03	
10	0	3	1	2	6	2	3	4	1	6	0.00	0.00	
11	2	0	6	5	1	0	0	3	0	5	0.00	0.00	
12	2	4	1	5	1	6	5	2	1	4	0.29	0.00	
13	1	4	4	5	2	2	0	4	5	3	0.00	0.00	
14	6	0	3	4	5	2	0	6	6	3	0.00	0.00	
15	0	0	0	4	2	6	2	4	5	0	0.00	0.00	
16	4	0	5	6	0	3	2	5	1	2	0.29	0.00	
17	6	5	4	4	4	4	4	3	4	2	0.40	0.02	
18	5	5	0	3	4	0	3	2	3	6	0.32	0.01	
19	5	1	6	5	2	3	2	3	2	3	0.32	0.01	
20	2	3	0	3	0	3	3	6	3	6	0.29	0.00	
21	1	0	3	4	6	3	0	5	3	1	0.00	0.00	
22	2	3	4	5	0	6	5	2	3	6	0.33	0.01	
23	0	4	1	6	6	3	3	5	1	4	0.00	0.00	
24	0	5	3	1	0	4	5	5	3	6	0.00	0.00	
25	2	2	1	3	2	4	1	2	1	4	0.19	0.00	
$Q^*(A_{ij})$	0.31	0.24	0.28	0.40	0.24	0.33	0.28	0.34	0.26	0.40	0.17	0.00	
$Q_e^*(B_i)$	0.31	0.28		0.35			0.28	0.34	0.35				
$Q^*(B_i)$	0.31	0.46		0.70			0.28	0.34	0.56				

Scale										
0	1	2	3	4	5	6	7	8	9	10
0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

Confidence intervals				
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e
0.90		0.12	0.22	
0.95	0.17	0.11	0.24	0.16
0.99		0.09	0.26	

Table 5: Checklist results, scores and confidence intervals - Example 2.

k	B_1		B_2		B_3			B_4	B_5	B_6		$Q_e^{*(k)}$	$Q^{*(k)}$
	A_{11}	A_{21}	A_{22}	A_{31}	A_{32}	A_{33}	A_{41}	A_{51}	A_{61}	A_{62}			
1	10	0	2	4	6	2	5	10	9	2	0.50	0.07	
2	2	9	3	4	10	0	4	5	6	3	0.49	0.03	
3	7	10	6	6	3	6	0	6	2	8	0.00	0.00	
4	3	3	5	3	4	3	9	6	8	9	0.52	0.07	
5	6	4	4	4	3	0	7	0	3	5	0.00	0.00	
6	7	5	3	0	9	7	5	5	1	3	0.47	0.04	
7	7	1	0	5	9	0	5	7	0	3	0.33	0.01	
8	10	1	2	9	4	3	2	1	3	4	0.30	0.00	
9	2	9	7	0	10	4	10	1	6	7	0.47	0.02	
10	1	5	10	10	6	0	2	0	4	4	0.00	0.00	
11	0	4	10	8	7	8	3	7	10	6	0.00	0.00	
12	9	2	9	6	6	10	4	2	10	0	0.61	0.07	
13	5	2	10	10	1	0	0	10	3	5	0.00	0.00	
14	2	5	10	9	9	8	0	0	0	4	0.00	0.00	
15	5	4	2	3	5	7	3	2	7	4	0.38	0.01	
16	8	2	3	2	7	9	1	6	3	2	0.36	0.01	
17	4	3	9	0	7	7	7	2	3	7	0.48	0.04	
18	7	4	0	6	5	8	10	2	10	9	0.52	0.05	
19	9	9	5	8	0	5	2	1	0	4	0.35	0.01	
20	6	10	3	5	4	9	3	10	3	10	0.71	0.17	
21	5	7	2	9	4	1	6	7	3	0	0.47	0.05	
22	4	10	10	8	1	8	6	2	7	8	0.54	0.04	
23	9	7	6	0	0	0	9	1	9	5	0.00	0.00	
24	9	1	4	1	6	7	6	0	6	10	0.00	0.00	
25	5	3	8	3	10	1	6	1	0	3	0.38	0.01	
$Q^*(A_{ij})$	0.57	0.48	0.53	0.49	0.54	0.45	0.46	0.38	0.46	0.50	0.31	0.03	
$Q_e^*(B_i)$	0.57	0.61		0.67			0.46	0.38	0.56				
$Q^*(B_i)$	0.57	0.74		0.89			0.46	0.38	0.71				

Scale										
0	1	2	3	4	5	6	7	8	9	10
0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

Confidence intervals				
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e
0.90	0.31	0.24	0.39	0.24
0.95		0.22	0.41	
0.99		0.19	0.44	

Table 6: Checklist results, scores and confidence intervals - Example 3.

Table 6 contains a corresponding sampling result. No grey value is dominating now, all values between dark grey and white are represented. The empirical evaluation scores of single goals range between 0.38 and 0.57 and reflect at first glance in tendency an average satisfaction level. The same applies to key goals. These scores ranging between 0.38 and 0.67. Nevertheless, the total empirical evaluation score $Q_e^*(C)$ amounts only 0.31 which is quite far away from an average satisfaction at a rate of 50%. Again this is a hint to the right direction. If we consider the scores $Q_e^{*(k)}(C)$ of checklist data records, then it appears that the total goal has been failed for eight times.

Hence the examples underline as well the advantages as the necessity of structure oriented evaluation. Non-observance of logical goal structure may cause misjudgement.

The wide confidence intervals as well as the comparatively high value of empirical standard deviation $\sigma_e = 0.24$ are a hint to a particularly pronounced random background during interrogation.

3.6 Summary

In this section a measure theoretical approach for evaluation of e-learning is developed. Via measure spaces for the involved e-learning components a product measure space is generated. This product space allows the calculation of scores for evaluation of logical structured goal structures of e-learning projects. The product measure satisfies the same calculation rules like any normalised measure of general measure theory.

Depending on contribution of e-learning components to achievement of an e-learning goal logical series and parallel structures play an important part. For these goal structures corresponding score calculation rules are derived. By combination of these rules a consistent evaluation of complex goal structures becomes possible.

Based on the calculation rules an estimation method for scores is developed. By this method empirical scores can be calculated for sampling results obtained by means of adapted checklists.

For a comparison of scores for evaluation of different e-learning processes a score calibration is recommended. This calibration transforms a

score value into an evaluation score value. This value enables a further interpretation of score values and is helpful at final analysis of sampling data for evaluation report. An estimation for the evaluation score based on checklist results can be obtained like at empirical score.

The precision of empirical scores or empirical evaluation scores can be estimated by confidence intervals.

Some simulation examples are considered which illustrate for selected evaluation situations the developed evaluation method and demonstrate the advantages of SURE model against linear evaluation models.

The presented evaluation method is by their embedding into the general measure theory quite universal applicable. In an analogous manner is an evaluation of efficiency of administration or organisation processes possible. A further application example is the evaluation of robustness of embedded systems during runtime. We refer to [55] and [18].

4 Pilot testing of SURE model in evaluation process

"...Evaluation means, literally, assigning a value to something. It answers questions about the value of your e-learning course or project. Evaluation teaches you about how well your learning solutions work and how you can improve them..." [21].

William Horton. (President of William Horton Consulting, Inc.)

In this section we consider some evaluation examples for courses of Mongolian University of Science and Technology (MUST) and Technische Universität Chemnitz (TUC). The examples refer to following courses:

1. "Application software for engineering computation". A web based e-learning course for Master students of MUST, code S.CS701.
2. "Maintenance of computers". Virtual laboratory for Bachelor students, Power Engineering School of MUST, code M.EC308.
3. "Instruction for gas station and liquid gas station". Virtual laboratory for Bachelor students, School of Geology and Petroleum Engineering of MUST, code C.PE535.
4. "Practice-Oriented Introduction to Computer Graphics". Video lecture for local and foreign students, Department of Computer Science, TUC.

4.1 Application software for engineering computation

The Mongolian University of Science and Technology is one of main universities for Mongolian engineering education. It is a multidisciplinary and multilevel university for education, training and scientific research. MUST is also the largest center for scientific and cultural exchanges in Mongolia.

Academic year	Term	E-Students	E-Courses
2007/2008	Spring	62	19
2008/2009	Autumn	81	8
2008/2009	Spring	59	13
2009/2010	Autumn	116	20
2009/2010	Spring	113	20
2010/2011	Autumn	251	23
2010/2011	Spring	447	43
2011/2012	Autumn	532	71
2011/2012	Spring	976	124
2012/2013	Autumn	1205	150
2012/2013	Spring	904	161
2013/2014	Autumn	1181	211

Table 7: Statistic of e-learning at MUST.

MUST is the leading university in Mongolia with respect to e-learning applications in higher education.

In 2007 the university started to offer web based distance learning courses for master students (see Table 7). The table shows that the interest in e-learning courses at MUST has increased very substantially since 2007.

MUST uses an own learning management system (LMS) which is called Unimis and by which all online courses are managed.

The evaluation mission of e-learning course S.CS701 is to assess quality of course and to identify those parts of e-learning which need improvements for next academic year, to evaluate quality of tutoring and to collect feedback from students to improve quality of content design.

Step 1. Definition of key goals

The evaluation team defined six key goals (see Table 8 and Figure 12a). These are:

- Course content;
- Content transfer methodology;

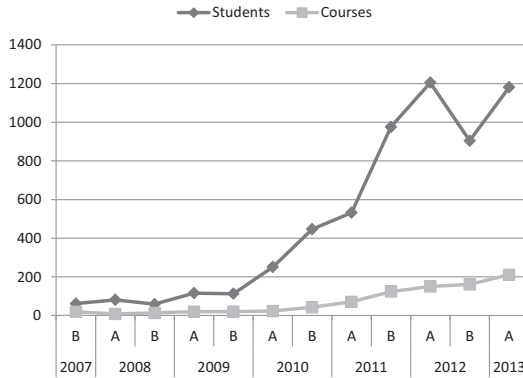


Figure 11: E-courses and students in e-learning of MUST.

- Tutor's skills;
- Virtual classroom environment;
- Course design;
- Learning management system.

Step 2. Definition of sub goals

In the actual case, the evaluation team described the key goals by sub goals, as it is shown in Figure 12b and Table 8.

The evaluation team has to define the sub goals carefully. If the evaluation team accepts only one way or method to reach a key goal then in sense of SURE model the key goal consists of only on sub goal which is identical with the key goal.

Step 3. Confirmation of evaluation goals

After first two steps logical structure of key and sub goals should be confirmed by evaluation team. The evaluation team has to check again all evaluation goals and should discuss once again the evaluation structure.

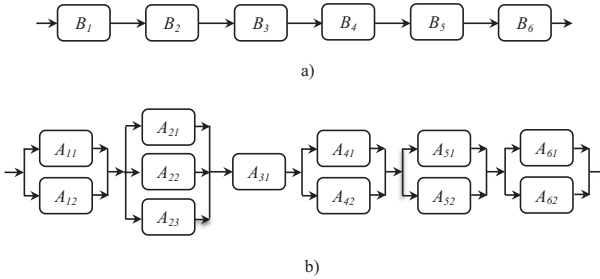


Figure 12: Logical structure of key (a) and sub (b) goals of S.CS701

When the members of evaluation team can agree with evaluation goals this should be fixed by a protocol.

Step 4. Creation of checklist

The checklist has to be adapted to logical structure of evaluation goal. The questions have to be clear, unambiguous and understandable. They should be formulated in short sentences. If special terms are used we have to explain these terms by a glossary. Figure 13 shows the corresponding proposal.

Step 5. Acceptance of checklist.

Before the final acceptance of checklist, the evaluation team should be given a further opportunity to discuss and modify the questions if necessary. This is a similar step like the confirmation of evaluation goal structure and prevents later conflicts. When the final form of checklist is established this should be fixed again by a protocol.

Step 6. Data collection

For data collection exist several possibilities. In any case one should guarantee anonymity. For the example considered here we have three samples. The results are displayed in Table 9, Table 10 and Table 11.

The first example, Table 9, refers to autumn term of 2011. The evaluation scale ranges from 1 up to 6, sample size is $n = 24$. The second

Key goal	Sub goal	Goals of evaluation
B_1	A_{11}	Usable of content
	A_{12}	Course content quality
B_2	A_{21}	Reading materials
	A_{22}	Audio materials
	A_{23}	Video materials
B_3	A_{31}	Tutor skills
B_4	A_{41}	Collaborative learning environment
	A_{42}	Virtual room environment
B_5	A_{51}	Added values of course design
	A_{52}	Satisfaction about course design
B_6	A_{61}	Registration process
	A_{62}	User interface of LMS

Table 8: The evaluation goals of S.CS701.

example, Table 10, refers to autumn term of 2012, evaluation interval was $[0..5]$, sample size is $n = 20$. The third example, Table 11, refers to autumn term of 2013, evaluation scale ranges from 1 up to 5, sample size is $n = 21$. The data were collected online.

Step 7. Data processing

Tables 9-11 contain the corresponding score estimation values. The estimation values for the evaluation scores are:

- $Q_e^*(C) = 0.84$ for sample of Autumn term of 2011;
- $Q_e^*(C) = 0.83$ for sample of Autumn term of 2012;
- $Q_e^*(C) = 0.89$ for sample of Autumn term of 2013.

Moreover the tables contain the following estimation values:

- Estimation values $Q^*(A_{ij})$ for the single scores $Q(A_{ij})$, $i = 1, \dots, r$, $j = 1, \dots, s_i$.
- Estimation values $Q_e^{*(k)}(C)$ for the scores $Q_e(C)$ for each checklist data record, $k = 1, \dots, n$, $n = 24, 20$ and 21.

Key	Sub	Statement	disagree - totally agree					
			0	1	2	3	4	5
B1	1	Course content						
	A11	Course content was useful to increase my knowledge						
	A12	Course content level fitted to offered course						
B2	2	Content transfer methodology						
	A21	Reading materials were supportive for study						
	A22	Audio materials were useful for study						
	A23	Video materials were helpful for study						
B3	3	Tutor's skills						
	A31	Tutors skills were excellent						
B4	4	Virtual classroom environment						
	A41	Exercises and communication tools were useful for learning						
	A42	Virtual learning environment was kindly and warm						
B5	5	Course design						
	A51	Multimedia effects in course design were perfect						
	A52	Course design was excellent						
B6	6	Learning management system						
	A61	Registration process was easy						
	A62	Learning management system worked well						

Figure 13: The checklist proposal.

- Estimation values $Q_e^*(B_i)$ for the evaluation scores $Q_e(B_i)$ of key goals B_i , $i = 1, \dots, r$.

Step 8. The evaluation report

These evaluations were self-assessment evaluations of teaching faculty. The main aim of these evaluations was to collect information with respect to course relevance, effectiveness, efficiency, and quality of courses to support development and upgrading of e-learning process.

During the three years of evaluation, all together 65 students participated in the survey. This are about 35% of all learners. All evaluations have been performed as summative evaluations at the end of the course.

Figure 14 shows all empirical evaluation scores are greater that 0.5 and the values are quite high. We can summarize here that these e-learning courses have been successfully. For a more detailed analysis we have to look in the empirical evaluation scores for key goals. In first year, Autumn

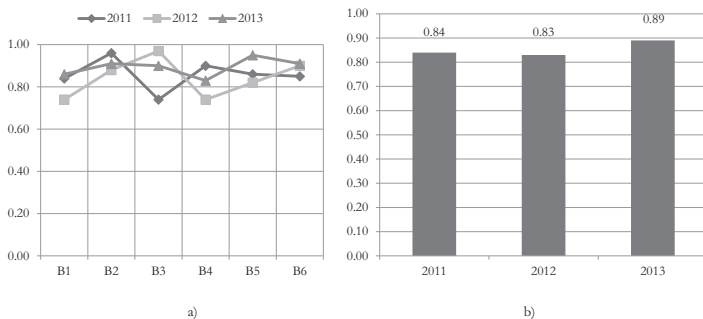


Figure 14: Evaluation scores of key goals.

term of 2011, key goal B_3 had the smallest score. The empirical evaluation score of goal B_3 is 0.74. The highest empirical evaluation score received key goal B_2 with a value of 0.96. In the next year, Autumn term of 2012, the lowest scores have been observed for key goals B_1 and B_4 with a value of 0.74 in both cases. Best empirical evaluation score key goal B_6 with a value 0.90. In last year, Autumn term of 2013 key goal B_4 processes lowest score. Evaluation score of this key goal is 0.83. Best empirical evaluation score has been obtained for key goal B_2 with a score of 0.93. Hence we can summarize that, two times key goal B_4 has been evaluated with lowest score by the learners. Therefore here we have to look back to the key goal structure.

The key goal B_4 refers to 'Virtual classroom environment' and consists of two sub goals. For further analysis we have to check how sub goals A_{41} and A_{42} achieved their goals. Sub goal A_{41} has been evaluated by 0.82, 0.73, and 0.70 during the three evaluation processes. The sub goal A_{42} has been evaluated by 0.84, 0.73 and 0.86. This shows sub goal A_{41} has been evaluated uniformly not better than sub goal A_{42} . Again we have to look back to sub goal structure and can check the corresponding questions. The question for sub goal A_{41} was formulated as: "Exercises and communication tools were useful for learning". That means, we should check exercises which are included into e-course materials and we should perhaps improve communication tools between learners in virtual learning environment.

k	B_1		B_2			B_3	B_4		B_5		B_6		$Q_e^{*(k)}$
	A_{11}	A_{12}	A_{21}	A_{22}	A_{23}	A_{31}	A_{41}	A_{42}	A_{51}	A_{52}	A_{61}	A_{62}	
1	4	4	4	5	6	3	4	5	4	2	4	3	0.58
2	4	5	6	6	4	4	6	5	4	5	4	6	0.82
3	4	4	5	6	5	4	3	6	6	5	4	6	0.84
4	4	6	5	6	6	5	6	4	5	4	5	5	0.88
5	6	5	5	6	6	5	4	4	5	6	5	6	0.88
6	4	4	5	3	3	2	5	4	4	4	4	4	0.51
7	6	5	6	6	6	5	5	6	6	6	5	6	0.96
8	4	5	6	5	6	4	5	4	5	5	5	6	0.79
9	6	4	6	6	6	6	4	6	6	6	4	6	1.00
10	5	6	6	6	5	6	6	5	4	4	5	5	0.88
11	6	4	6	6	6	6	5	6	6	6	4	6	1.00
12	6	6	6	6	6	6	6	6	5	6	5	6	1.00
13	4	5	5	4	6	6	4	5	6	6	4	6	0.90
14	6	5	6	4	4	4	5	6	4	5	4	6	0.87
15	4	5	6	6	6	6	5	6	4	6	2	4	0.82
16	6	6	5	5	3	4	6	6	5	4	3	5	0.76
17	4	4	6	6	6	5	4	5	5	5	3	6	0.81
18	5	5	6	6	4	4	6	5	5	5	5	6	0.85
19	5	4	6	6	6	5	5	5	6	6	4	6	0.88
20	5	6	6	5	3	6	6	6	5	4	5	6	0.95
21	4	5	5	6	6	3	4	5	6	6	3	5	0.72
22	4	6	5	5	4	5	6	6	5	6	3	5	0.86
23	4	4	6	6	6	4	6	3	5	6	4	5	0.80
24	6	6	6	4	5	5	6	6	5	6	3	5	0.90
$Q^*(A_{ij})$	0.77	0.79	0.92	0.88	0.83	0.74	0.82	0.84	0.81	0.83	0.61	0.88	0.84
$Q_e^*(B_i)$	0.84		0.96			0.74	0.90		0.86		0.85		

Confidence intervals					Scale					
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e	1	2	3	4	5	6
0.90	0.84	0.80	0.88	0.12	0.00	0.20	0.40	0.60	0.80	1.00
0.95		0.80	0.89							
0.99		0.78	0.91							

Table 9: Evaluation scores for Autumn term of 2011 (S.CS701).

k	B_1		B_2			B_3	B_4		B_5		B_6		$Q_e^{*(k)}$
	A_{11}	A_{12}	A_{21}	A_{22}	A_{23}	A_{31}	A_{41}	A_{42}	A_{51}	A_{52}	A_{61}	A_{62}	
1	5	4	5	5	5	5	4	5	5	4	5	5	1.00
2	2	1	3	4	4	5	3	2	1	3	3	4	0.58
3	4	4	4	4	4	5	4	4	4	4	4	4	0.83
4	3	4	4	4	4	5	4	4	4	4	4	4	0.82
5	3	2	2	3	2	5	2	2	3	2	2	4	0.56
6	4	5	5	4	5	5	4	4	5	4	5	5	0.96
7	3	3	5	4	5	5	4	4	4	4	5	5	0.85
8	2	3	3	5	4	5	4	4	4	3	5	3	0.81
9	2	3	4	5	3	5	2	2	5	5	5	5	0.77
10	5	4	4	5	4	5	5	5	5	5	5	5	1.00
11	3	3	5	4	4	4	5	5	4	4	5	5	0.85
12	5	5	5	5	5	5	3	3	4	3	3	4	0.82
13	3	2	3	4	3	5	1	3	4	4	3	5	0.70
14	4	5	5	4	4	5	3	2	4	4	4	4	0.83
15	3	5	5	5	5	5	3	4	5	5	5	5	0.95
16	4	4	4	4	4	5	5	5	5	5	5	5	0.93
17	4	4	2	3	3	4	4	4	4	3	4	4	0.74
18	3	4	5	5	5	5	4	4	4	5	4	5	0.91
19	4	4	4	4	4	5	5	5	4	4	5	4	0.89
20	3	4	5	5	5	4	4	2	4	4	4	4	0.79
$Q^*(A_{ij})$	0.69	0.73	0.82	0.86	0.82	0.97	0.73	0.73	0.82	0.79	0.85	0.89	0.83
$Q_e^*(B_i)$	0.74		0.88			0.97	0.74		0.82		0.90		

Confidence intervals				Scale						
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e	0	1	2	3	4	5
0.90		0.79	0.87		0.00	0.20	0.40	0.60	0.80	1.00
0.95	0.83	0.78	0.88	0.12						
0.99		0.76	0.90							

Table 10: Evaluation scores for Autumn term of 2012 (S.CS701).

k	B_1		B_2			B_3	B_4		B_5		B_6		$Q_e^{*(k)}$
	A_{11}	A_{12}	A_{21}	A_{22}	A_{23}	A_{31}	A_{41}	A_{42}	A_{51}	A_{52}	A_{61}	A_{62}	
1	4	3	5	4	4	5	4	5	4	5	5	5	0.95
2	5	5	5	3	3	5	4	5	5	5	5	5	1.00
3	4	4	5	2	2	4	4	4	5	5	5	5	0.89
4	5	4	5	4	3	5	4	5	5	5	5	5	1.00
5	5	5	5	3	3	5	1	3	5	4	4	4	0.84
6	4	4	4	5	4	4	4	5	4	5	5	3	0.93
7	5	3	5	3	2	5	4	5	5	5	5	5	1.00
8	4	4	3	4	4	5	4	4	5	5	4	5	0.88
9	4	5	4	3	3	5	4	5	5	5	5	3	0.94
10	4	3	4	3	3	4	1	3	5	4	4	4	0.72
11	5	4	5	4	4	5	4	5	5	5	4	4	0.96
12	2	3	2	2	2	2	1	3	3	3	3	2	0.47
13	4	3	5	4	5	4	4	5	5	4	5	5	0.91
14	5	5	5	3	3	5	5	5	4	5	4	5	1.00
15	4	4	5	2	2	4	4	4	3	2	3	4	0.76
16	5	4	5	5	4	5	4	5	4	5	4	5	1.00
17	4	4	5	3	3	5	1	3	5	4	4	4	0.81
18	4	4	4	5	5	4	4	3	4	5	5	3	0.88
19	5	3	5	3	2	5	4	4	5	4	5	5	0.96
20	4	4	3	5	4	5	4	4	5	5	4	5	0.93
21	4	5	4	3	5	4	4	5	5	5	5	3	0.96
$Q_e^*(A_{ij})$	0.86	0.79	0.89	0.70	0.67	0.90	0.70	0.86	0.91	0.90	0.89	0.85	0.89
$Q_e^*(B_i)$	0.87		0.93			0.90	0.83		0.96		0.93		

Confidence intervals				Scale						
Level	$Q_e^*(C)$	$q_{e,0}^*$	$q_{e,1}^*$	σ_e	0	1	2	3	4	5
0.90		0.85	0.94		0.00	0.20	0.40	0.60	0.80	1.00
0.95	0.89	0.84	0.95	0.13						
0.99		0.82	0.97							

Table 11: Evaluation scores for Autumn term of 2013 (S.CS701).

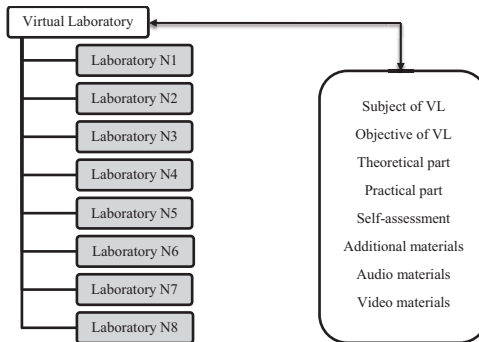


Figure 15: Basic standard for virtual laboratories.

4.2 Virtual laboratory for maintenance of computers

Engineering education is cost-intensively. Modern laboratories with good technical and technological equipment are the base for high quality in teaching and learning at engineering universities. The recent fast and dynamical development of technology requires continuous upgrading of laboratories what frequently is connected with financial problems. A way out of this situation is to create and use virtual laboratories in teaching. Virtual laboratories can be used in e-learning and at traditional class room teaching. An essential advantage of virtual laboratories is that they are easy to upgrade to latest technological standard.

MUST developed several virtual laboratories and these laboratories are included into teaching process. One of these laboratories is the virtual laboratory for bachelor students: "Maintenance of computers, M.EC308".

The Faculty defined the basic requirements for virtual laboratories, see Figure 15. Based on this standard the virtual laboratories of PES are created.

During Spring term of 2011, for the first time, virtual laboratories have been used in traditional face to face teaching.

The evaluation mission of this test was to identify teaching relevant aspects for the further development of virtual laboratories by evaluation by different involved groups.

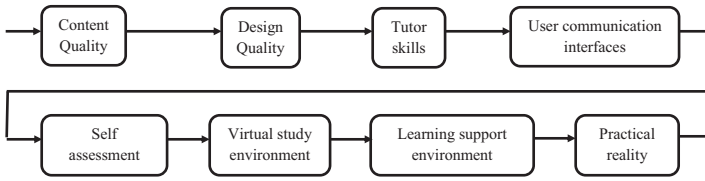


Figure 16: Key goals of evaluation for virtual laboratory M.EC308.

Step 1. Definition of key goals

By the evaluation team eight key goals have been defined for evaluation. The logical structure of key goals is shown by Figure.16. All eight dimension are important to describe the quality of a virtual laboratory.

- Content quality
- Design quality
- Tutor skills
- User communication interfaces
- Self assessment
- Virtual study environment
- Learning support environment
- Practical reality

Step 2. Definition of sub goals

Beside the key goals the evaluation team defined several sub goals for the key goals B_1 , B_2 , B_3 , B_6 and B_8 , see Table 12.

B_i	A_{ij}	Sub goal
B_1	A_{11}	Content clearness
	A_{12}	Content level
	A_{13}	Content design
B_2	A_{21}	Design of usability
	A_{22}	Design of virtual laboratory
B_3	A_{31}	Tutor skills to support virtual environment
	A_{32}	Tutor feedback during study
B_4	A_{41}	User interface
B_5	A_{51}	Self assessment
B_6	A_{61}	Collaborative learning environment
	A_{62}	Virtual room for learners
B_7	A_{71}	Introduction information
	A_{72}	Usefulness of practice
	A_{71}	Tutor and learner cooperation
B_8	A_{81}	Virtual reality quality

Table 12: Confirmed evaluation goals.

Step 3. Confirmation of evaluation goals

The evaluation team confirmed the defined evaluation goal structure by protocol.

Step 4. Creation of checklist

Figure 17 shows the checklist draft for the defined logical structure of evaluation goal.

Step 5. Acceptance of checklist

The checklist draft was discussed by the members of evaluation team. The evaluation team accepted checklist and confirmed that by a protocol.

Step 6. Data collection

Data have been collected in paper based version by means of checklist forms.

Key	Sub	Questions	disagree				agree	
			0	1	2	3	4	5
Content quality								
B1	A11	Content objectives was clearly defined						
	A12	Content level met my expectation						
	A13	Content was designed well						
Design quality								
B2	A21	Design of usability was well						
	A22	Virtual laboratory general design was good						
Tutor skill								
B3	A31	Tutors was supportive in virtual learning environment						
	A32	Tutors feedback during study was helpful						
User communication interface								
B4	A41	User communication interface was designed easy to use and understand						
	Self assessment quality							
B5	A51	Self assessment was good organized and well related to content						
	Virtual study environment							
B6	A61	Collaborative learning environment was useful to study						
	A62	Virtual room for learners was helpful during study						
Learning support environment								
B7	A71	Introduction had well organized full information						
	A72	Practical part of virtual laboratory was supportive for study						
	A73	Learning support environment had well organized tools between tutor and learner						
Practical reality								
B8	A81	Practical reality was developed well nearly to real environment						

Figure 17: Checklist for data collection.

Step 7. Data processing

Table 13 contains the corresponding evaluation scores. It holds

$$Q_e^*(C) = 0.82.$$

Moreover the tables contain the following score estimation values:

- Estimation values $Q^*(A_{ij})$ for the single scores $Q(A_{ij})$, $i = 1, \dots, r$, $j = 1, \dots, s_i$.
- Estimation values $Q_e^{*(k)}(C)$ for the score $Q_e(C)$ for each check list data record, $k = 1, \dots, n$, $n = 20$.
- Estimation values $Q_e^*(B_i)$ for the evaluation scores $Q_e(B_i)$ of key goals B_i , $i = 1, \dots, r$.

k	B ₁			B ₂		B ₃		B ₄	B ₅	B ₆		B ₇			B ₈	Q _e *
	A ₁₁	A ₁₂	A ₁₃	A ₂₁	A ₂₂	A ₃₁	A ₃₂	A ₄₁	A ₅₁	A ₆₁	A ₆₂	A ₇₁	A ₇₂	A ₇₃	A ₈₁	
1	4	5	4	3	4	4	3	3	4	4	3	5	5	4	4	0.78
2	3	5	4	4	5	3	4	5	4	4	3	4	4	4	4	0.85
3	4	5	5	4	4	4	5	4	5	4	3	4	4	4	3	0.83
4	3	3	4	4	4	4	5	5	4	4	4	4	5	4	4	0.85
5	4	3	4	3	4	4	4	4	3	3	4	4	4	4	3	0.72
6	4	4	4	3	4	4	5	5	4	4	5	5	5	4	5	0.91
7	5	5	4	4	4	5	5	5	5	5	4	5	5	4	4	0.95
8	4	5	5	5	5	4	4	4	5	5	5	4	4	4	5	0.92
9	3	4	4	3	3	3	4	4	4	3	3	3	5	5	4	0.75
10	4	3	3	5	5	4	4	5	4	4	5	3	5	5	3	0.85
11	3	5	5	4	4	4	3	3	4	3	4	4	5	5	4	0.79
12	4	4	5	3	3	5	4	4	3	3	4	5	5	4	4	0.80
13	3	3	2	2	3	3	2	2	3	3	4	3	3	4	2	0.53
14	2	4	3	5	4	3	3	3	4	5	4	4	3	3	4	0.75
15	4	3	4	5	2	4	3	4	3	4	4	4	4	4	4	0.78
16	4	4	5	5	4	4	4	4	5	5	3	3	4	5	4	0.92
17	4	4	4	5	5	5	4	4	5	3	5	4	5	5	4	0.92
18	3	5	4	5	4	5	5	4	5	4	5	4	5	4	5	0.97
19	5	4	5	5	4	4	4	4	5	5	4	3	4	3	4	0.88
20	5	4	3	4	5	4	3	2	3	4	5	3	4	3	3	0.72
	0.75	0.82	0.81	0.81	0.80	0.80	0.78	0.78	0.82	0.79	0.81	0.78	0.88	0.82	0.77	0.82
		0.87		0.85		0.82	0.78	0.82		0.86		0.89		0.77		

Confidence intervals					Scale					
Level	Q _e *	q _{e,0} *	q _{e,1} *	σ _e	0	1	2	3	4	5
0.90	0.82	0.79	0.86	0.10	0.00	0.20	0.40	0.60	0.80	1.00
0.95		0.78	0.87							
0.99		0.76	0.88							

Table 13: The evaluation scores for M.EC₃₀₈.

Step 8. The evaluation report

Main aim of evaluation was to confirm expectation of stakeholders on efficiency of a virtual laboratory and to collect observation data with respect to satisfactions of students with a virtual laboratory in comparison with traditional teaching. The evaluation score for M.EC₃₀₈ was 0.82. This value reflects an average satisfaction of learners with the virtual laboratory at a rate of 82%. This evaluation result met the expectations of stakeholders.

Additionally we can analyse how the aims of key goals have been reached. The top evaluation score for a key goal was obtained by key goal B₇, the

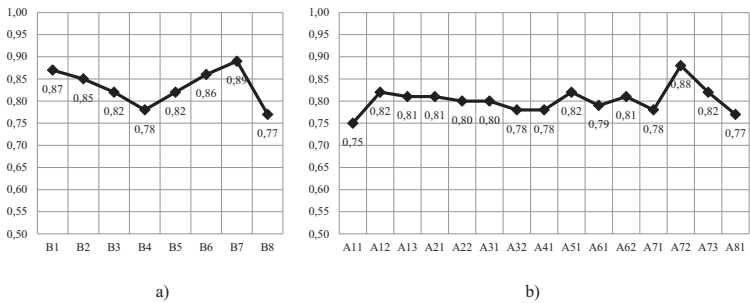


Figure 18: Evaluation scores for virtual laboratory M.EC308.

worst one by key goal B_8 . An empirical score of 0.89 for key goal B_7 shows that learners have been very satisfied with the results for key goal "Learning support environment". In contrast to that the evaluation of results of B_7 - "Practical reality" - with a rate of 77% are a hint that one should think about on improvements of virtual laboratory in this direction.

Figure 18 emphasises graphically the evaluation results of key and sub goals. The highest ranked sub goal is A_{72} . The students confirm quite clear that the practical part of virtual laboratory was supportive for study. The high value for the sub target A_{72} implies the high value for key target B_7 . This confirms that students seem to like the use of virtual laboratories in learning where the practical part of virtual laboratory is an especially estimated component of virtual laboratory.

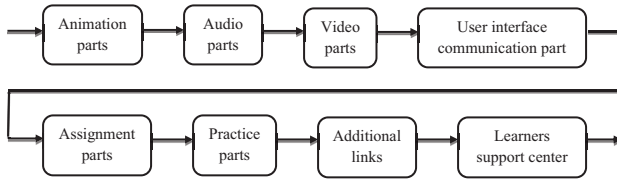


Figure 19: The logical structure of key goals for virtual laboratory C.PE353.

4.3 Construction gas station and liquid gas station

This is a further example for a virtual laboratory: "Construction gas station and liquid gas station, C.PE353". This laboratory was developed in cooperation of Power Engineering School with School of Geology and Petroleum Engineering of MUST.

The laboratory has been offered for first time during spring term in 2012 as a face to face unit at School of Geology and Petroleum Engineering. The evaluation mission of this test was to get a first impression how this laboratory works from the perspective of involved groups.

Step 1. Definition of key goals

For this evaluation eight key goals have been defined. The evaluation team selected goals which should reflect to developers how the single parts of laboratory have been perceived among involved groups. The following parts have been considered.

- Animation part
- Audio part
- Video part
- User interface communication part
- Assignment part
- Practice part

Targets	Parts	not important					very important					
		0	1	2	3	4	5	6	7	8	9	10
B1	animation parts											
B2	audio parts											
B3	video parts											
B4	UIC parts											
B5	assignment parts											
B6	practice parts											
B7	additional links parts											
B8	learners support center											

Figure 20: Accepted checklist for C.PE353.

- Additional links part
- Learners support center

Step 2. Definition of sub goals

The mission of this evaluation was to analyse expectations and satisfaction with laboratory of different groups which have been interested or have been involved into this laboratory. For this reason the evaluation team decided not to define sub goals.

Step 3. Confirmation of evaluation goals

The logical structure of evaluation goals was confirmed by evaluation team and fixed by a protocol.

Step 4. Creation of checklist

The checklist has been created in accordance with defined goal structure, see Figure 19.

Step 5. Acceptance of checklist

The draft of checklist was confirmed by evaluation team and fixed by protocol.

Step 6. Data collection

Data collection was performed paper based. The sample size of learners

k	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	$Q_e^{*(k)}(C)$
	A_{11}	A_{21}	A_{31}	A_{41}	A_{51}	A_{61}	A_{71}	A_{81}	
1	5	5	9	7	7	8	5	5	0.62
2	9	7	9	9	9	9	8	9	0.86
3	9	9	8	8	9	10	7	10	0.87
4	8	8	8	8	8	8	8	8	0.80
5	6	7	3	8	10	10	8	6	0.68
6	8	10	8	8	9	9	10	5	0.82
7	6	7	3	8	10	10	8	6	0.68
8	9	7	9	9	9	9	8	9	0.86
9	10	8	10	9	10	9	10	10	0.95
10	10	9	8	8	9	9	9	8	0.87
$Q^*(A_{ij})$	0.80	0.77	0.75	0.82	0.90	0.91	0.81	0.76	0.80
$Q_e^*(B_i)$	0.80	0.77	0.75	0.82	0.90	0.91	0.81	0.76	

Scale										
0	1	2	3	4	5	6	7	8	9	10
0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

Table 14: The evaluation scores of Learners.

was not so high, because of this laboratory is offered during last term of Bachelor study where students have the possibility to select this course as major study course.

Step 7. Data processing

The calculation of sample scores has been performed according the rules of Section 3. The Tables 14 until 16 contain the corresponding empirical evaluation scores. The estimation value for evaluation score $Q_e(C)$ for learners is $Q_e^*(C) = 0.80$, for stakeholders $Q_e^*(C) = 0.72$ and for developers $Q_e^*(C) = 0.81$.

Moreover the tables contain estimation values $Q_e^{*(k)}(C)$ for evaluation score $Q_e^*(C)$ for each checklist result, for the single evaluation scores $Q(A_{ij})$ as well estimation values for evaluation scores of key goals $Q_e^*(B_i)$. The estimation values for key goals are once again summarized in Table 17 depending on considered groups.

k	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	$Q_e^{*(k)}(C)$
	A_{11}	A_{21}	A_{31}	A_{41}	A_{51}	A_{61}	A_{71}	A_{81}	
1	10	6	9	10	7	10	9	9	0.86
2	5	4	4	5	4	10	10	6	0.56
3	9	7	5	6	8	9	9	7	0.74
$Q^*(A_{ij})$	0.80	0.57	0.60	0.70	0.63	0.97	0.93	0.73	0.72
$Q_e^*(B_i)$	0.80	0.57	0.60	0.70	0.63	0.97	0.93	0.73	

Table 15: Evaluation scores of Stakeholders.

k	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	$Q_e^{*(k)}(C)$
	A_{11}	A_{21}	A_{31}	A_{41}	A_{51}	A_{61}	A_{71}	A_{81}	
1	8	5	8	8	10	10	9	9	0.82
2	8	8	8	8	8	8	8	8	0.80
$Q^*(A_{ij})$	0.80	0.65	0.80	0.80	0.90	0.90	0.85	0.85	0.81
$Q_e^*(B_i)$	0.80	0.65	0.80	0.80	0.90	0.90	0.85	0.85	

Table 16: Evaluation scores of Developers.

Step 8. The evaluation report

Main target of this evaluation was to find out how the involved groups are satisfied with actual version of virtual laboratory.

The virtual laboratory is welcome by learners. Students are satisfied. This is confirmed by the quite high empirical evaluation score $Q_e^*(C) = 0.8$. The key goals are evaluated by students quite homogeneously. That means, for further improvement in sense of students we have to go with our questionnaire more into the details.

Similar tendencies we observe for the other both groups which have been involved into evaluation. All empirical evaluation scores are higher than 0.5. This hints to an over-average satisfactions at all groups. The sample sizes are quite small such that the observed differences between estimation values should be considered as random effects. The somewhat smaller value which has been obtained from stakeholders should kept in mind for future evaluations.

Figure 21 presents a radar graphic which contains the evaluation results for the eight key goals of involved groups. This graphic underlines once again the homogeneity of evaluation.

Key goals	Learners	Stakeholders	Developers
Animation	0.80	0.80	0.80
Audio	0.77	0.57	0.65
Video	0.75	0.60	0.80
User interface	0.82	0.70	0.80
Assignments	0.90	0.63	0.90
Practices	0.91	0.97	0.90
Additional links	0.81	0.93	0.85
Learners support tools	0.76	0.73	0.85

Table 17: General evaluation scores of all three involved groups.

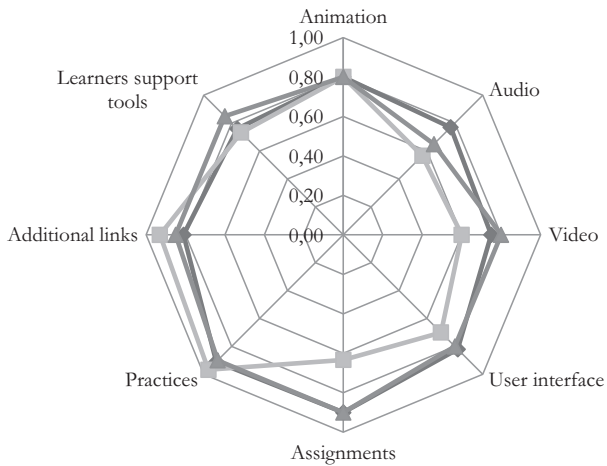


Figure 21: Radar graphic of evaluation results.

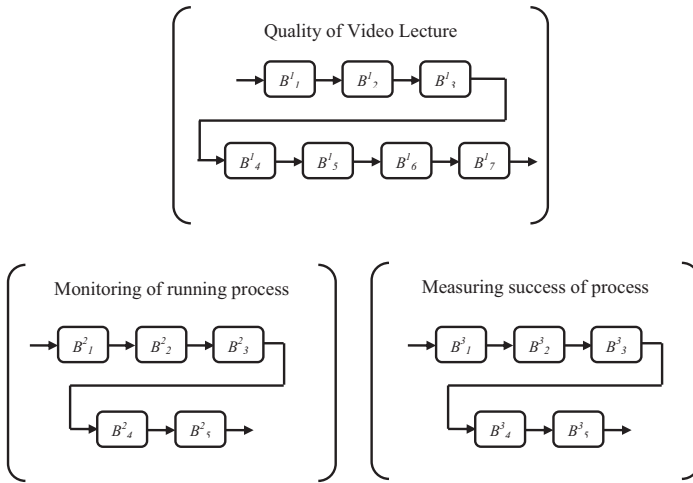


Figure 22: Space of different logical structures.

4.4 Practice-oriented introduction to computer graphics

Step 1. Definition of key goals

The Technische Universität Chemnitz started to develop an 'Online Platt form für akademisches Lehren und Lernen' (OPAL) in September 2005. The online platform OPAL bases on the open source learning platform "Online Learning And Training" (OLAT) which has been developed by Zürich University in 1999. Since 2007 the Computer Science Department of TU Chemnitz created several e-courses in English language and is offering distance learning courses for local students and students from abroad. One of the first distance learning courses was the video lecture "Practice-Oriented Introduction to Computer Graphics".

To define key goals evaluation team cooperated with stakeholders well. The evaluation team and stakeholders had a series of discussions. During these discussions it became clear that the stakeholders had different expectations on the evaluation process. Finally, three different goal structures have

Key goals	Sub goals	Indicators
$B_1^{(1)}$	$A_{11}^{(1)}$	Information
	$A_{12}^{(1)}$	Payment process
$B_2^{(1)}$	$A_{21}^{(1)}$	Web design
	$A_{22}^{(1)}$	User interface
	$A_{23}^{(1)}$	Topic design
$B_3^{(1)}$	$A_{31}^{(1)}$	Content quality
	$A_{32}^{(1)}$	Content objective
	$A_{33}^{(1)}$	Learning materials
	$A_{34}^{(1)}$	Recording quality
$B_4^{(1)}$	$A_{41}^{(1)}$	Tutor skills
	$A_{42}^{(1)}$	Tutor support
$B_5^{(1)}$	$A_{51}^{(1)}$	Level of exercises
	$A_{52}^{(1)}$	Usefulness and supportive
$B_6^{(1)}$	$A_{61}^{(1)}$	Learners collaboration environment
	$A_{62}^{(1)}$	Usefulness of collaboration
$B_7^{(1)}$	$A_{71}^{(1)}$	Course level
	$A_{72}^{(1)}$	Knowledge
	$A_{73}^{(1)}$	Skill ability
	$A_{74}^{(1)}$	Course fee
	$A_{75}^{(1)}$	General satisfaction

Table 18: Key and sub goals for "Quality of Video lecture".

been confirmed.

First logical structure of key goals $B_1^{(1)}, \dots, B_7^{(1)}$ was defined for evaluation of 'Quality of Video lecture':

- Administration process
- OPAL system
- Course content
- Tutoring
- Exercise
- Collaboration

Key goals	Sub goals	Indicators
$B_1^{(2)}$	$A_{11}^{(2)}$ $A_{12}^{(2)}$	Information Payment process
$B_2^{(2)}$	$A_{21}^{(2)}$ $A_{22}^{(2)}$	Technical environments User and navigation environments
$B_3^{(2)}$	$A_{31}^{(2)}$ $A_{32}^{(2)}$	Schedule for exercises Learning style
$B_4^{(2)}$	$A_{41}^{(2)}$ $A_{42}^{(2)}$	Self tests in lectures Self test useful
$B_5^{(2)}$	$A_{51}^{(2)}$	Assignment and exams

Table 19: Key and sub goals for "Monitoring of process".

- Satisfactions

Second logical structure of key goals $B_1^{(2)}, \dots, B_5^{(2)}$ was defined for evaluation of 'Monitoring of running process':

- Administration process
- OPAL system
- Learning activity of learners
- Self motivation
- Course assessments

Third logical structure of key goals $B_1^{(3)}, \dots, B_5^{(3)}$ was defined for evaluation of 'Success of running process':

- Administration process
- Learning part
- Video Lecture

Key goals	Sub goals	Indicators
$B_1^{(3)}$	$A_{11}^{(3)}$ $A_{12}^{(3)}$	Registration Payment
$B_2^{(3)}$	$A_{21}^{(3)}$ $A_{22}^{(3)}$	Learning materials Exercises
$B_3^{(3)}$	$A_{31}^{(3)}$ $A_{32}^{(3)}$ $A_{33}^{(3)}$ $A_{34}^{(3)}$	Audio Learning style Language Time
$B_4^{(3)}$	$A_{41}^{(3)}$ $A_{42}^{(3)}$	Usage Access
$B_5^{(3)}$	$A_{51}^{(3)}$ $A_{52}^{(3)}$	Tutor skills Tutor support

Table 20: Key and sub goals for "Measuring success of learning process".

- Online library
- Teaching

The lists and logical structures of key goals show some overlaps. Data were collected by the same questionnaire and data processed according to the three considered logical structures.

Step 2. Definition of sub goals

Evaluation team defined sub goals for each key goal structure of evaluation. Table 18 shows the key and sub goals for "Quality of Video lecture", Table 19 the key and sub goals for "Monitoring of process" and Table 20 the key and sub goals for "Measuring success of learning process".

Step 3. Confirmation of evaluation goals

After definition of key and sub goals, the logical structures and checklist were checked and reviewed by stakeholders.

No	Question	Impact	Comment
1	Information about goals of course was clear	Agree	
2	Information about advantages of course was clear	Agree	
3	Information about course duration was clear	Totally agree	
4	Information about course schedule was clear	— Choose —	
5	Information about course exams was clear	— Choose —	
6	Information about course fee was clear	Totally agree Agree Neutral Disagree	
7	Course fee payment process had convenient possibilities of money transfer	Totally disagree Not applicable	
8	I received confirmation of payment in good time	— Choose —	
9	Registration process was clear to start	— Choose —	
10	Opportunity to use online library was good for my self motivation and learning	— Choose —	

Figure 23: Online checklist design view.

Step 4. Creation of checklist

For this example we used a beta version of tool for SURE model to design and create checklist. Each question has a drop down selection list. This list includes following choices:

- Totally agree = 5;
- Agree = 4;
- Neutral = 3;
- Disagree = 2;
- Totally disagree = 1;
- Not applicable = 0.

The questions have moreover an additional a comment field. If the user selects 'Not applicable' as answer, the comment field is activated and the user can enter there a comment. If an user selected as answer 'Not applicable' we completed this answer by average over remaining answers.

Step 5. Acceptance of checklist

Before confirmation of checklist evaluation team and involved groups tested online survey. After that the checklist was accepted.

Step 6. Data collection

Data were collected based on the confirmed checklists by online survey. Each learner received unique access code to the online survey by email. That guarantees the anonymity of the survey. Forty six learners attended to this distance learning since 2010, nine participants responded to the survey.

Step 7. Data processing

The Tables 21, 22 and 23 contain the corresponding sampling results and score estimation values.

The estimation values for the evaluation score:

- $Q_e^*(C) = 0.79$ for Quality of Video lecture;
- $Q_e^*(C) = 0.74$ for Monitoring of running learning process;
- $Q_e^*(C) = 0.75$ for Success of running process.

Moreover the tables contain the following estimation values:

- Estimation values $Q^*(A_{ij})$ for the single scores $Q(A_{ij})$, $i = 1, \dots, r$, $j = 1, \dots, s_i$.
- Estimation values $Q_e^{*(k)}(C)$ for the scores $Q_e^{(k)}(C)$ for each checklist data record, $k = 1, \dots, 9$.
- Estimation values $Q_e^*(B_i)$ for the evaluation scores $Q_e(B_i)$ of key goals B_i , $i = 1, \dots, r$.

Step 8. The evaluation report

This evaluation had several advantages compared with the previous evaluations:

- Three different logical structures have been considered simultaneously which were adapted to the different expectations of stakeholders on evaluation.

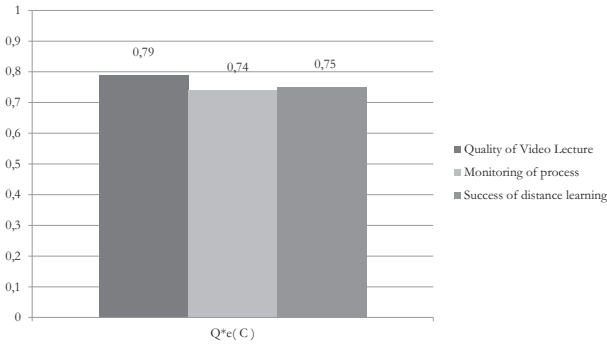


Figure 24: Evaluation results for three logical structures.

- The data collection was performed as online survey by means of the SURE model tool based on a common checklist.

Figure 24 shows a summary of evaluation results. All evaluation scores are higher than 0.5. This indicates an over average satisfaction for all three groups where Quality of Video Lecture was evaluated highest with an empirical evaluation score of $Q^*_e(C) = 0.79$. This reflects an average satisfaction of users at a rate of 79%.

Based on the evaluation results, it can be summarized that the distance learning course "Practice-Oriented Introduction to Computer Graphics" runs successfully. However, we have to note that sample sizes are quite small such that these results are to consider as orientation.

k	B_1		B_2			B_3				B_4	
	A_{11}	A_{12}	A_{21}	A_{22}	A_{23}	A_{31}	A_{32}	A_{33}	A_{34}	A_{41}	A_{42}
1	4	3	4	3	5	5	5	5	5	4	5
2	4	3	4	3	4	4	3	3	4	4	4
3	5	3	2	2	5	4	4	5	5	3	3
4	4	3	4	4	4	3	4	4	4	4	4
5	4	3	3	2	3	4	4	4	5	3	3
6	4	5	4	4	4	4	4	4	4	4	4
7	5	2	4	4	5	4	5	4	4	3	2
8	4	4	4	4	4	4	4	3	4	4	4
9	5	5	5	5	4	5	5	5	5	5	5
$Q^*(A_{ij})$	0.83	0.61	0.69	0.61	0.81	0.78	0.81	0.78	0.86	0.69	0.69
$Q_e^*(B_i)$	0.82		0.82			0.87				0.71	

k	B_5		B_6		B_7					$Q_e^{*(k)}(C)$
	A_{51}	A_{52}	A_{61}	A_{62}	A_{71}	A_{72}	A_{73}	A_{74}	A_{75}	
1	4	5	5	5	4	5	4	4	4	0.94
2	4	4	4	4	4	4	4	3	4	0.70
3	4	4	3	3	5	4	4	4	5	0.79
4	3	4	3	4	5	5	4	4	4	0.73
5	4	4	3	3	4	5	4	4	4	0.66
6	4	4	5	5	5	5	4	4	4	0.85
7	3	3	2	3	5	5	5	5	5	0.69
8	4	4	4	4	3	4	3	4	4	0.73
9	4	5	5	5	5	5	5	5	5	1.00
$Q^*(A_{ij})$	0.69	0.78	0.69	0.75	0.86	0.92	0.78	0.78	0.83	0.79
$Q_e^*(B_i)$	0.77		0.73		0.93					

Scale				
1	2	3	4	5
0.00	0.25	0.50	0.75	1.00

Table 21: Evaluation scores for "Quality of Video lecture".

k	B_1		B_2		B_3		B_4		B_5	$Q_e^{*(k)}(C)$
	A_{11}	A_{12}	A_{21}	A_{22}	A_{31}	A_{32}	A_{41}	A_{42}	A_{51}	
1	5	4	4	4	5	3	4	4	5	0.89
2	4	3	4	3	3	3	3	3	3	0.55
3	4	3	3	4	4	4	3	4	5	0.73
4	4	3	4	3	3	4	4	4	4	0.69
5	4	3	3	3	4	5	3	3	3	0.60
6	3	5	4	4	5	4	4	4	4	0.84
7	5	1	5	4	4	4	1	3	3	0.64
8	3	4	3	4	4	4	3	4	4	0.69
9	5	5	5	5	5	4	4	5	5	1.00
$Q^*(A_{ij})$	0.78	0.61	0.72	0.69	0.78	0.72	0.56	0.69	0.75	0.74
$Q_e^*(B_i)$	0.80		0.73		0.82		0.65		0.75	

Table 22: Evaluation scores for "Monitoring of process".

k	B_1		B_2		B_3				B_4		B_5		$Q_e^{*(k)}(C)$
	A_{11}	A_{12}	A_{21}	A_{22}	A_{31}	A_{32}	A_{33}	A_{34}	A_{41}	A_{42}	A_{51}	A_{52}	
1	5	5	5	5	5	4	5	5	5	5	5	5	1.00
2	4	3	3	4	4	4	3	4	4	4	3	4	0.68
3	5	5	5	4	5	4	4	4	2	2	4	4	0.72
4	4	4	4	4	4	4	4	4	4	3	5	3	0.77
5	4	3	3	4	5	5	4	4	3	3	3	3	0.64
6	5	4	4	4	4	3	4	4	3	4	4	4	0.76
7	3	3	5	3	4	4	4	3	4	3	1	4	0.65
8	5	4	4	4	4	4	4	4	3	2	3	3	0.64
9	5	5	5	5	5	5	4	5	4	4	5	4	0.94
$Q^*(A_{ij})$	0.86	0.75	0.81	0.78	0.86	0.78	0.75	0.78	0.64	0.58	0.67	0.69	0.75
$Q_e^*(B_i)$	0.84		0.84		0.85			0.62		0.74			

Table 23: Evaluation scores for "Measuring success of learning process".

4.5 Summary

This section summarizes results of pilot examples for the SURE model and includes four test evaluations which were performed using the SURE model.

Section 4.1 describes the evaluation of master course CS701 of MUST. Main aim of this evaluation was to measure the quality of the e-content, tutor skills and satisfaction of learners in collaborative learning environment. For this aim six key goals have been considered. Checklist was designed as questionnaire. We collected data from the last three academic years: Autumn terms of 2011, 2012 and 2013. The empirical evaluation scores of these evaluations were 0.84, 0.83 and 0.89. These values confirm that this e-learning course runs successfully and stable.

Section 4.2 refers to the evaluation of a virtual laboratory of MUST. Aim of this evaluation for virtual laboratory M.EC308 was an estimation of the quality of this virtual laboratory. Checklist was designed again as a questionnaire, the data have been collected by a paper based version. Evaluation score was 0.82. This results means that this project is running successfully.

Section 4.3 describes the evaluation of a virtual laboratory, C.PS353. The main target of this evaluation was to identify the interests and needs of different groups which are involved in development and implementation of virtual laboratory. Three groups have been considered: Learners, Stakeholder and Developers. By the learners the virtual laboratory was evaluated with 0.80, by stakeholders with 0.72 and by developers with 0.82.

Section 4.4 describes the evaluation of a video lecture of TU Chemnitz. This video lecture is a distance learning course which is offered by TU Chemnitz. For evaluation have been considered three different logical goal structures. The checklist was organized as general checklist and data were collected by the beta version of SURE tool.

The evaluation scores for all three logical structures were higher than 0.5 which indicates an above average satisfaction. The users evaluated 'Quality of video lecture' with $Q_e^*(C) = 0.79$. That means average satisfaction of users for quality of video lecture was 79%. For 'Monitoring of distance learning' we obtained the evaluation score $Q_e^*(C) = 0.74$ and for 'Measurement of Success of learning' an evaluation score $Q_e^*(C) = 0.75$.

These empirical results confirm that distance learning course "Practice-Oriented Introduction to Computer Graphics" runs successfully. An over average level of satisfaction in sense of all considered goal structures has been observed. We have to note that sample size have been small here such that we should consider these results as orientation.

5 The research results

In frame of this study a structure oriented evaluation model with measure theoretical based score calculation rules for evaluation of e-learning is developed.

An e-learning is a complex process which involves many different groups with specific tasks and roles. To evaluate such a process it should be possible to include into evaluation the expectations and interests of all involved groups. Moreover, the evaluation method has to be able to take reference to the logical structure of total goal of an e-learning. Finally, the corresponding evaluation principles and their score calculation rules should be consistent with the calculation rules of general measure theory which forms the proven scientific basis for measurements in geometrical context or the evaluation of probabilities in probability and statistics, for instants.

This is realized by the structure oriented evaluation model (SURE model) developed here. The model consists of eight steps which form the methodical concept for a structure oriented evaluation.

An essential advantage of SURE model is that this model gets along without of weight factors as they are needed in linear evaluation models. Hence the SURE model is more objective than linear models.

For the evaluation of a logical goal structure we distinguish between key and sub goals which are linked by logical AND and OR relations. Key goals are goals which have to be achieved such that the main goal can be achieved. Sub goals describe different goal alternatives to achieve a key goal or the main goal.

By means of measure spaces in sense of general measure theory for the sub goals and embedding of the relevant sub goals into a product measure space a multidimensional evaluation of complex goal structures becomes possible.

The obtained product space allows the calculation of scores for the evaluation of logical structured goals of e-learning. The product measure satisfies the same calculation rules like any normalised measure of general measure theory.

Depending on the contribution of sub goals to achievement of an e-learning goal logical series and parallel structures play an important part.

For these goal structures corresponding score calculation rules are derived. By combination of these rules a consistent evaluation of complex goal structures becomes possible.

Based on the calculation rules an estimation method for scores is developed. By this method empirical scores can be calculated for sampling results obtained by adapted checklists.

For a comparison of scores for evaluation of different e-learning processes a score calibration is recommended. This calibration transforms a SURE model score value into an evaluation score value. This value enables a further interpretation of score values and is helpful at final analysis of sampling data for evaluation report. An estimation for the evaluation score based on checklist results can be obtained like at empirical score. The precision of empirical scores or empirical evaluation scores can be estimated by confidence intervals.

Some simulation and pilot examples are considered which illustrate for selected evaluation situations the developed evaluation method.

The presented evaluation method is by their embedding into the general measure theory quite universal applicable. In an analogous manner is an evaluation of efficiency of administration or organisation processes possible. A further application example is the evaluation of robustness of embedded systems during runtime. We refer to [56] and [18].

Our further investigations are directed to the following topics:

- Further improvement of the beta version of web based application of SURE model.
- Development of a cloud computing server for on-line processing of checklist data in sense of SURE model.
- Extension of statistical part of SURE model for further data analysis of checklist data.
- Exploitation of further applications for the SURE model in context with evaluation of complex systems or processes with logical oriented goal structures.

Recently, the SURE model is used for evaluation processes at Mongolian University of Science and Technology, Ulaanbaatar, Mongolia, and at Tech-

nische Universität Chemnitz, Chemnitz, Germany. It is intended to integrate the SURE model into Unimis and OPAL system of the two universities.

List of Figures

1	Evaluation and e-learning framework.	21
2	Types and visualization of evaluation goals in SURE model.	29
3	The different types of logical structures.	30
4	Algorithm for logical structure of key goals.	31
5	The structure oriented evaluation model.	33
6	The logical structures of evaluation goals.	34
7	The waterfall diagram for SURE model.	40
8	The architecture of tool for SURE model.	41
9	Series and parallel process.	47
10	Goal structures for simulation examples.	68
11	E-courses and students in e-learning of MUST.	79
12	Logical structure of key (a) and sub (b) goals of S.CS701	80
13	The checklist proposal.	82
14	Evaluation scores of key goals.	83
15	Basic standard for virtual laboratories.	87
16	Key goals of evaluation for virtual laboratory M.EC308.	88
17	Checklist for data collection.	90
18	Evaluation scores for virtual laboratory M.EC308.	92
19	The logical structure of key goals for virtual laboratory C.PE353.	93
20	Accepted checklist for C.PE353.	94
21	Radar graphic of evaluation results.	97
22	Space of different logical structures.	98
23	Online checklist design view.	102
24	Evaluation results for three logical strucrues.	104

List of Tables

1	List of evaluation models.	17
2	Google search engine result.	25
3	Checklist design proposal.	36
4	Checklist results, scores and confidence intervals - Example 1.	70
5	Checklist results, scores and confidence intervals - Example 2.	72
6	Checklist results, scores and confidence intervals - Example 3.	73
7	Statistic of e-learning at MUST.	78
8	The evaluation goals of S.CS701.	81
9	Evaluation scores for Autumn term of 2011 (S.CS701). . . .	84
10	Evaluation scores for Autumn term of 2012 (S.CS701). . . .	85
11	Evaluation scores for Autumn term of 2013 (S.CS701). . . .	86
12	Confirmed evaluation goals.	89
13	The evaluation scores for M.EC308.	91
14	The evaluation scores of Learners.	95
15	Evaluation scores of Stakeholders.	96
16	Evaluation scores of Developers.	96
17	General evaluation scores of all three involved groups. . . .	97
18	Key and sub goals for "Quality of Video lecture".	99
19	Key and sub goals for "Monitoring of process".	100
20	Key and sub goals for "Measuring success of learning process".	101
21	Evaluation scores for "Quality of Video lecture".	105
22	Evaluation scores for "Monitoring of process".	106
23	Evaluation scores for "Measuring success of learning process".	106

References

- [1] Alkin, M. C. (1969). "Evaluation Theory Development". *Evaluation Comment*, No.2. pp.2-7.
- [2] Alkin, M. C. and Fitz-Gibbon, C. T. (1975). "Methods and Theories of Evaluating Programs", *Journal of Research and Development in Education*, Vol.8. No.3. pp.2-15.
- [3] Alkin, M. C. and Christie, C. A. (2004). *An Evaluation Theory Tree*. SAGE Press. pp.12-65.
- [4] Allen, I. E. and Seaman, J. (2011). "Going the Distance Online Education in the United States, 2011", *Babson Survey Research Group and Quahog Research Group, LLC*, pp.8-16. [Online]. Available: <http://www.onlinelearningsurvey.com/reports/goingthedistance.pdf>.
- [5] Attwell, G. Ed., (2006). *Evaluating e-learning A guide to the evaluation of e-learning*, Evaluate Europe Handbook Series, Vol.2. Perspektiven-Offset-Druck, Bremen, Germany.
- [6] Bauer, H. (2001). *Measure and Integration Theory*. De Gruyter Studies in Mathematics, Berlin.
- [7] Betsy, M. A. et al., (2005). "WeBAL: A Web-Based Assessment Library to Enhance Teaching and Learning in Engineering", *IEEE Transactions on Education*, Vol.48. No.4. pp.764-771.
- [8] Black, P. and Wiliam, D. (1998). "Assessment and classroom learning", *Assessment in Education*, Vol.5. No.1. pp.7-74.
- [9] Blayney, P. and Freeman, P. (2004). "Automated formative feedback and summative assessment using individualised spreadsheet assignments", *Australasian Journal of Educational Technology*, Vol.20. No.2. pp.209-231. [Online]. Available: <http://www.ascilite.org.au/ajet/ajet20/blayney.htm>
- [10] Bridge, P. and Appleyard, A. (2008). "A comparison of electronic and paper-based assignment submission and feedback", *British Journal of Educational Technology*, Vol.39. No.4. pp.644-650. [Online]. Available: <http://www.pgce.soton.ac.uk/ict/NewPGCE/PDFs/A>
- [11] Colace, F., De Santo, M. and Vento, M. (2005). "E-learning platform: Developing an Evaluation, Strategy in a Real Case", *Frontiers in Education 35th Annual Conference*. pp.20-25.
- [12] Colace, F., De Santo, M. and Pietrosanto, A. (2006.) "Evaluation Models for E-Learning Platform: an AHP approach", *36th ASEE/IEEE Frontiers in Education Conference*, pp.1-6.

- [13] Ehlers, U. D. (2005). *Quality in e-learning: use and dissemination of quality approaches in European e-learning: a study by the European Quality Observatory*, Office for Official Publications of the European Communities.
- [14] Ehlers, U. D. and Pawlowski, J. M. (Eds.), (2006). *Handbook of Quality and Standardisation in E-Learning*, Springer Berlin - Heidelberg.
- [15] Ehlers, U. D. (2010). *Low Cost, Community Based Certification for E-learning in Capacity Building*, Bonn.
- [16] Eisner, E. W. (1975). "The perceptive eye: Toward the reformation of educational evaluation", American Educational Research Association, Washington, DC.
- [17] *E-learning: Industry Fact Book*, April/May Elearning 2013, Vol.9. No.3. pp.28-29. [Online]. Available: <http://elmezine.epubxp.com/i/133739/10>.
- [18] Heller, A. (2013). *Systemeigenschaft Robustheit - Ein Ansatz zur Bewertung und Maximierung von Robustheit eingebetteter Systeme*, PhD Thesis, Win. Schriftenreihe 'Eingebettete, Selbstorganisierende Systeme', Vol.12. Universitätsverlag Chemnitz.
- [19] Hogan, L. R. (2007). "The Historical Development of Program Evaluation: Exploring the Past and Present", *Online Journal of Workforce Education and Development*, Vol.2. Iss.4. pp.1-14.
- [20] Hogo, M. A. (2010). "Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study", *International Journal of Computer Science and Information Security (IJCSIS)*, Vol.7. No.2. pp.131-140. [Online]. Available: <http://arxiv.org/ftp/arxiv/papers/1003/1003.1499.pdf>.
- [21] Horton, W. (2001). *Evaluating E-learning*. Alexandria, VA: ASTD.
- [22] *How to Perform Evaluations - Evaluation reports*, (2002). Canadian International Development Agency. [Online]. Available: <http://www.oecd.org/derec/canada/35138852.pdf>.
- [23] John, B. and Phyllis, D. *The ABSs of Evaluation*, (2005). Jossey-Bass.
- [24] Khan, B. H. (2004). "The People, process and product continuum in e-learning: The e-learning P₃ model", *Educational Technology*, Vol.44. No.5. pp.33-40. [Online]. Available: <http://asianvu.com/bookstoread/etp/elearning-p3model.pdf>
- [25] Khan, B. H. (2005). *Managing e-learning: Design, delivery, implementation and evaluation*. Hershey, PA: Information Science Publishing.
- [26] Khan, B. H. (2006). *Flexible Learning in an Information Society*. Hershey PA: IGI Global.
- [27] Kirkpatrick, D. L. (1959). "Techniques for evaluating training programs", *Journal of American Society of Training Directors*, Vol.13. No.3. pp.21-26.

- [28] Kirkpatrick, D. L. (1994). *Evaluating Training Programs*. San Francisco: Berrett-Koehler Publishers, Inc.
- [29] Kirkpatrick, D. L. and Kirkpatrick, J. D. (2006). *Evaluating Training Programs. The Four Levels (3rd Edition)*. San Francisco.
- [30] Kaufman, R., Keller, J. and Watkins, R. (2007). "What works and what doesn't: Evaluation beyond Kirkpatrick", *Performance and Instruction*, Vol.35. No.2. pp.8-12. [Online]. Available:<http://home.gwu.edu/~rwatkins/articles/whatwork.PDF>
- [31] Lam, P. and McNaught, C. (2008). "A Three-Layered Cyclic Model of E-Learning Development and Evaluation", *Journal of Interactive Learning Research*, Vol.19. No.2. pp.313-329.
- [32] Lukas, L. (2014). *E-Learning Consulting LLC*, [Online]. Available: <http://www.e-learningconsulting.com/consulting/what/e-learning.html>.
- [33] Messick, S. J. (1986). *The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement*, ETS Research Report. RR-86-30.
- [34] Messick, S. J. (1994). "The Interplay of Evidence and Consequences in the Validation of Performance Assessments", *Educational Researcher*, Vol.23. No.2. pp.13-23. [Online]. Available: <http://edr.sagepub.com/content/23/2/13.full.pdf>
- [35] Patton, M. Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills, CA: Sage Publication.
- [36] Patton, M. Q. (2008). *Utilization-focused evaluation*. Sage Publications, Inc.
- [37] Penny, B. B. (2004). "An Online Assistant for Remote, Distributed Critiquing of Electronically Submitted Assessment", *Educational Technology and Society*, Vol.7. No.1. pp.29-41.
- [38] Phillips, J. J. (1991). *Handbook of Evaluation and Measurement Methods*, London: Gulf Press.
- [39] Phillips, J. J. (1997). *Return on Investment in Training and Performance Improvement Programs*, Houston: Gulf Publishing Co.
- [40] Phillips, P. P. and Phillips, J. J. (2008). *ROI Fundamentals: Why and When to Measure Return on Investment (Measurement and Evaluation Series)*, San Francisco, CA: Pfeiffer.
- [41] Phillips, P. P. et al., (2010). *ASTD Handbook of Measuring and evaluating training*. Alexandria, VA: ASTD.
- [42] *E-learning quality, Aspects and criteria for evaluation of e-learning in higher education*, (2008). Hoegskoleverket, Swedish National Agency for Higher Education, Report 2008:11 R.

- [43] Ruhe, V and Zumbo, B.D. (2009). *Evaluation in distance education and e-learning: the unfolding model*. New York: The Guilford press.
- [44] Scriven, M. S. et. al., (1967). "The methodology of evaluation". *Perspectives of Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation, No.1 pp.39-81. Chicago: Rand McNally.
- [45] Scriven, M. S. (1991). "Pros and cons about goal-free evaluation", *American Journal of Evaluation*, Vol.12. No.1. pp.55-76.
- [46] Scriven, M. S. (1999). "The nature of evaluation part II: Training", *Practical Assessment, Research and Evaluation*, Vol.6. No.12. pp.1-5.
- [47] Scriven, M. S. (2012). *The three revolutions*, [Online]. Available:<http://michaelscriven.info/papersandpublications.html>.
- [48] Stufflebeam, D. L. (1983). "The CIPP model for Program Evaluation", *Evaluation in Education and Human Services*, Vol.6. pp.117-141.
- [49] Stufflebeam, D. L. (2000). "The CIPP Model for Evaluation", *Evaluation in Education and Human Services*, Vol.49. pp.279-317.
- [50] Stockley, D. (2014). *The Training, Learning and Performance Management*. Consultant Derek Stockley Pty Ltd. [Online]. Available: <http://derekstockley.com.au/elearning-definition.html>.
- [51] Taylor-Powell, E. and Henert, E. (2008). *Developing a logic model: Teaching and training guide*. Board of Regents of the University of Wisconsin System. Program Development and Evaluation, Madison. [Online]. Available: <http://www.uwex.edu/ces/pdande/evaluation/pdf/lmguidecomplete.pdf>.
- [52] Tudevdagva, U. and Hardt, W. (2011). *A new evaluation model for e-learning programs*, Technical Report CSR-11-03, Chemnitz University of Technology.
- [53] Tudevdagva, U. and Hardt, W. (2012). "A measure theoretical evaluation model for e-learning programs", in *Proceedings of the IADIS on e-Society*, Berlin, Germany, pp.44-52.
- [54] Tudevdagva, U. et al., (2012). "New Approach for E-Learning Evaluation", in *Proceedings of the 7th International Forum on Strategic Technology*, Tomsk, Russia, pp.712-715.
- [55] Tudevdagva, U., Heller, A. and Hardt, W. (2013). "A model for robustness evaluation of embedded systems", in *Proceedings of the IFOST conference*, Ulaanbaatar, Mongolia. pp.288-292.
- [56] Tyler, R. W. (1949). "Basic principles of curriculum and instruction", *Chicago: The University of Chicago Press*.
- [57] Tyler, R. W. (1972). "General statement of evaluation", *The Journal of Educational Research*, Vol.35. No.4. pp.492-501.

- [58] UNESCO. (2011). *Global Education Digest 2011 - Comparing Education Statistics Across the World*, UNESCO Institute for Statistics, Montreal: Canada.
- [59] Warr, P., Bird, M. and Rackham, N. (1970). *Evaluation of Management Training: A Practical Framework with Cases, for Evaluating Training Needs and Results*. Aldershot: Gower Publishing.
- [60] Zaharias, P. and Poulymenakou, A. (2009). "Developing a usability evaluation method for e-learning applications: beyond functional usability", *International Journal of Human-Computer Interaction*, Vol.25. No.1. pp.75-98.
- [61] Zhang, W. and Cheng, Y.L. (2012). "Quality Assurance in E-Learning: PDPP Evaluation Model and its Application", *The international review of research in open and distance learning*, Vol.13. No.3. pp.66-82. [Online]. Available: <http://files.eric.ed.gov/fulltext/EJ1001012.pdf>

Related materials

- [62] Buzzetto-More, N. A. and Alade, A. J. (2006). *Best Practices in e-Assessment*. Journal of Information Technology Education, Vol.5. pp.251-269. [Online]. Available: <http://informingcience.org/jite/documents/Vol5/v5p251-269Buzzetto152.pdf>.
- [63] Barbara, M. et. al, (2010). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. U.S. Department of Education, Washington, D.C.
- [64] Campbell, A. (2005). "Application of ICT and rubrics to the assessment process where professional judgement is involved: the features of an e-marking tool", *Assessment and Evaluation in Higher Education*, Vol.30. No.5. pp.529-537.
- [65] Cronbach, L. J. et al., (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*, San Francisco: Jossey-Bass.
- [66] Cronbach, L.J. (1982). *Designing Evaluations of Educational and Social Programs*, San Francisco: Jossey-Bass Inc Pub.
- [67] Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*, Thousand Oaks, CA: Sage.
- [68] Edwards, K. I. et al., (2002). "EAST: developing an electronic assessment and storage tool", *Assessment and Evaluation in Higher Education*, Vol.27. No.1. pp.95-104.
- [69] Frydenberg, J. (2002). *Quality Standards in e-Learning: A matrix of analysis*, Irvine Distance Learning Center, University of California. [Online]. Available: <http://www.irrodl.org/index.php/irrodl/article/view/109/189>.

- [70] Guba, E. G. and Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*, San Francisco: Jossey-Bass.
- [71] Geetha, U. K., Sherlyne, A. and Tudevtagva, U. (2012). "Intercultural understanding among Asian participants in a blended e-learning course", in *Proceedings of the First International conference on open and distance e-learning*, Manila, Philippines. pp.101-111.
- [72] Grif, M. G., Tudevtagva, U. and Tsoy, E. B. (2004). "Man Machine Systems Design Methods", *Science and Technology, KORUS*, Proceedings. The 8th Russian-Korean International Symposium, Vol. 3. pp.42-45.
- [73] Joint Committee on Standards for Educational Evaluation, (1994). *Program Evaluation Standards : How to Assess Evaluations of Educational Programs*, SAGE Publications, Inc.
- [74] Joseph, S. W. (1979). *Evaluation: Promise and Performance*, Washington: DC, Urban Institute Press.
- [75] Judith, B. S. (2002). "An Assessment of the Effectiveness of e-learning in Corporate Training Programs", *The international review of research in open and distance learning*, Vol.3. No.1. [Online]. Available: <http://www.irrodl.org/index.php/irrodl/article/view/83/16>.
- [76] Köhler, T., Neumann, J. and Saupé, V. (2010). *Online-Lernen. Ein Handbuch für das Lernen mit Internet*, München, Oldenbourg Wissenschaftsverlag.
- [77] Köhler, T., Kahnwald, N. and Reitmaier, M. (2008). *Lehren und Lernen mit Multimedia und Internet*, Medienpsychologie: Berlin: Springer.
- [78] Madaus, G. G., Scriven, M. S. and Stufflebeam, D. L. (1984). "Educational evaluation and accountability: A review of quality assurance efforts", *The American*, Vol.27. No.5. pp.649-673.
- [79] Master Plan for E-learning of Mongolian University of Science and Technology, (2007). *Mongolian University of Science and Technology*, [Online]. Available: <http://must.edu.mn/beta3/menu312>.
- [80] Mandinach, E. B. (2005). "The Development of Effective Evaluation Methods for E-Learning: A Concept Paper and Action Plan", *Teachers College Record*, Vol.107. No.8. pp.1814-1835, [Online]. Available: <http://www.csupomona.edu/dolce/pdf/mandinach.pdf>.
- [81] Oliver, M. (2000). *An introduction to the evaluation of learning technology*. Educational Technology and Society, Vol.3. No.4. pp.20-30.
- [82] Paul, F. M. (1997). *The Logic Model for Program Planning and Evaluation*, University of Idaho, [Online]. Available: <http://www.uiweb.uidaho.edu/extension/LogicModel.pdf>.

- [83] Rossi, P. H. and Wright, S. R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- [84] Shee, D. Y. and Wang, Y. S. (2008). *Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications*. Computers and Education, Vol.50. No.3. pp.894-905.
- [85] Stufflebeam, D. L. (1969). *A depth study of the evaluation requirement*, Theory Into Practice, Vol.5. No.3. pp.121-133.
- [86] Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision making*. Evaluation Center, Ohio State University.
- [87] Stufflebeam, D. L. and Webster, W. J. (1980). *An analysis of alternative approaches to evaluation*. Educational Evaluation and Policy Analysis, Vol.2. No.3. pp.5-20.
- [88] Stufflebeam, D. L. (1994). *Empowerment Evaluation, Objectivist Evaluation, and Evaluation Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go*. American Journal of Evaluation, No.15. pp.321-338.
- [89] Tudevtagva, U. and Batshagai, B. (2010). *Research and usage of virtual laboratory in teaching*. Proceedings of International Conference of 50th Anniversary of Power Engineering School of MUST, Ulaanbaatar, pp.26-31.
- [90] Tudevtagva, U. (2011). *Virtual Laboratories for e-Learning in Engineering Sciences Research and Implementation*, Poster paper, Faculty for the Future Forum, July, Paris, France.
- [91] Tudevtagva, U. and Sambuu, U. (2012). "A structure oriented evaluation model for readiness of egovernance". in *Proceedings of the 6th ICEGOV 2012*, Albany NY, USA, p.454-456.
- [92] Tudevtagva U., Hardt, W. and Ymchmaa, A. (2013). "E-education bridge between Europe and Asia-Tele Lecture". in *Proceedings of the e-CASE & e-Tech 2013*, Kitakyushu, Japan. pp.98-105.
- [93] Tudevtagva, U., Hardt, W. and Jargalmaa, D. (2013). "The development of logical structures for e-learning evaluation". in *Proceedings of the IADIS on e-Learning 2013*, Prag, Czech Republik, pp.431-435.
- [94] Tudevtagva, U. and Garamkhand, S. (2013). "Distance online course for librarian in Mongolia, reflection and learned lesson". in *Proceedings of the IADIS on e-Learning 2013*, Prag, Czech Republic, pp.483-486.
- [95] Weiss, C. H. (1999). "The Interface Between Evaluation and Public Policy", *The International Journal of Theory, Research, and Practice*, Vol.5 No.4. pp.468-486.

Web sites

- [96] Technische Universität Chemnitz, *Web site of distance course Tele-Lecture* [Online]. Available: <http://www.tu-chemnitz.de/informatik/iccs/index-modul1.php>
- [97] SURE model tool, *Online application tool for SURE model*, 2013, In Mongolian, [Online]. Available: <http://202.5.195.20/sure/default>
- [98] SURE model tool, *Online application tool for SURE model*, 2013, In English, [Online]. Available: <http://202.5.195.20/sure/eng/default>
- [99] Survey Software 1, *Survey Monkey*, [Online]. Available: <https://de.surveymonkey.com>
- [100] Survey Software 2, *Fluid Surveys*, [Online]. Available: <http://fluidsurveys.com/>
- [101] Survey Software 3, *Iperceptions*, [Online]. Available: <http://signup.iperceptions.com/online-survey.html?gclid=CPqAg63g-LgCFYKS3god1GAAUg>
- [102] Survey Software 4, *Free online surveys*, [Online]. Available: <http://freeonline-surveys.com/>
- [103] Survey Software 5, *Kwik Survey*, [Online]. Available: <http://kwiksurveys.com/>
- [104] Survey Software 6, *Easy polls*, [Online]. Available: <http://www.easypolls.net/>
- [105] Survey Software 7, *Survey planet*, [Online]. Available: <https://www.surveyplanet.com/>
- [106] Survey Software 8, *Sogo survey*, [Online]. Available: <http://www.sogosurvey.com/>
- [107] Survey Software 9, *Esurveyspro*, [Online]. Available: <http://www.esurveyspro.com/free-online-survey.aspx>
- [108] Survey Software 10, *Esurvey creator*, [Online]. Available: <https://www.esurveycreator.com>
- [109] Survey Software 11, *Stellar survey*, [Online]. Available: <http://stellarsurvey.com/>
- [110] Survey Software 12, *Questionpro*, [Online]. Available: <http://www.questionpro.com/>
- [111] Survey Software 13, *Esurv*, [Online]. Available: <http://esurv.org/>
- [112] Survey Software 14, *Questionform*, [Online]. Available: <http://questionform.com/>

- [113] Survey Software 15, *Panelplace*, [Online]. Available: <http://www.panelplace.com/>
- [114] Survey Software 16, *Surveycrest*, [Online]. Available: <http://www.surveycrest.com/>
- [115] Survey Software 17, *Addpoll*, [Online]. Available: <http://www.addpoll.com/>
- [116] Survey Software 18, *Quick Surveys*, [Online]. Available: <https://www.quick-surveys.com/>