

Assessing learning outcomes in two information retrieval learning environments

Kai Halttunen* & Kalervo Järvelin

Department of Information Studies, University of Tampere

FIN-33014 University of Tampere, Finland

Tel.: +358-3-215 8921

Fax.: +358-3-215 6560

E-mail:kai.halttunen@uta.fi, kalervo.jarvelin@uta.fi

[Preprint]

[Information Processing and Management 41(2005):4, 949-972]

* Corresponding author. Tel.: +358-3-215 8921 ; fax: +358-3-2156560. E-mail addresses:
kai.halttunen@uta.fi, kalervo.jarvelin@uta.fi

Abstract

In order to design information retrieval (IR) learning environments and instruction, it is important to explore learning outcomes of different pedagogical solutions. Learning outcomes have seldom been evaluated in IR instruction. The particular focus of this study is the assessment of learning outcomes in an experimental, but naturalistic, learning environment compared to more traditional instruction. The 57 participants of an introductory course on IR were selected for this study, and the analysis illustrates their learning outcomes regarding both conceptual change and development of IR skill. Concept mapping of student essays was used to analyze conceptual change and log files of search exercises provided data for performance assessment. Students in the experimental learning environment changed their conceptions more regarding linguistic aspects of IR and paid more emphasis on planning and management of search process. Performance assessment indicates that anchored instruction and scaffolding with an instructional tool, the IR Game, with performance feedback enables students to construct queries with fewer semantic knowledge errors also in operational IR systems.

Keywords: Learning outcomes; Information retrieval instruction; Learning environments; Conceptual change; Performance assessment

1 Introduction

Development of information retrieval (IR) tools and techniques have made IR a commonplace activity in different work, educational and leisure activities. End user searchers are now afforded with tools and have access to a wide variety of information sources that professional intermediaries had the sole right to use a decade ago. Information access skills have gained attention both in professional and all-round education.

Professional education within librarianship and information work has covered principles and use of IR systems since early online systems in the 70's and 80's. The invention of CD-ROM technology and OPAC's brought IR activities visible to end-users and brought about an intensive period of library and bibliographic instruction. The emerging of WWW technologies afforded users with a wide variety of information sources and tools to interact with from mid-1990s onwards.

Although professional intermediaries and end users of information retrieval systems have been educated since the advent of commercial online systems, and principles of IR systems belong essentially to curricula in Information Science, there are very few studies describing either pedagogical solutions used in instruction or reporting on assessment of learning outcomes (exceptions are Jacobson and Ignacio (1997) and Borgman, Gilliland-Swetland et al. (2001)). An review of literature both in education and information studies revealed that research on the assessment of learning outcomes is very rare in IR instruction. This holds for all levels of instruction: professional education of IR experts and intermediaries, bibliographic/user education, information skills and information literacy, and IR in different domains (like education, medicine, chemistry, and journalism). Assessment of learning outcomes is mainly based on surveys and self evaluation (see for example bibliographies by Rader (2000), Johnson (2001) and Johnson (2002)). The field of IR instruction is lacking a

systematic approach to learning outcome evaluation as well as methodologies for assessment (Pausch & Popp, 1997; Rader, 2002). The evaluation of learning outcomes, not just learning experiences, should direct the development of IR instruction and learning environments at different levels as pointed out earlier. It might also inform the development of IR systems as well.

This article, however, evaluates and compares between first year university students' learning outcomes in two IR learning environments. Looking at the outcomes is part of a larger research effort in order to develop and evaluate modules of an IR learning environment. The research project consists of four studies, namely:

- The design of Information Retrieval Game (also called Query Performance Analyzer), one module of the learning environment, and pilot evaluation of instructional use of this software application. (Halttunen & Sormunen, 2000; Sormunen, Halttunen, & Keskustalo, 2002; Sormunen et al., 1998)
- A study on students' prior conceptions of IR and their implications for the design of IR learning environments (Halttunen 2003a).
- An investigation on students' learning experiences and performance in two learning environments (Halttunen 2003b).
- An evaluation of learning outcomes. Findings of this investigation are reported in the present article.

In order to evaluate the instructional design and its experienced and observed effects on performance and learning outcomes, a design experiment was carried out. Tutored exercises were carried out in two different ways. In the *traditional learning environment*, different operational search systems were used to demonstrate the basic functions of IR systems. In

the *experimental setting*, full-text newspaper articles from local newspaper along with a press image database were used. These sources were used through the IR Game, a system which offers feedback to the searcher on the effectiveness of queries based on recall-base. The IR Game was based on the innovation to apply a test collection created for laboratory-based IR experiments as an instructional tool. Standard test collections offer a mechanism to generate search exercises with performance feedback since they provide a large set of documented search tasks (topics) and reliable estimates of the relevant documents available for each search task. The IR Game can be used to demonstrate and test the effectiveness of any individually formulated query for a search task in textual and image databases.

(Halttunen & Sormunen, 2000; Sormunen et al., 2002.) Various ideas of scaffolding and anchored instruction were applied in the experimental learning environment. Scaffolding refers to different kinds of supports that learners receive in their interaction with teachers, tutors and different kinds of tools within a learning environment as they develop new skills, concepts or levels of understanding. (Wood, Bruner & Ross, 1976; Gutzdial, 1994). The aim of the anchored instruction is to build semantically rich "anchors" that illustrate important problem solving situations. These anchors create a context that provides a common ground for experts as well as teachers and students from diverse backgrounds, to communicate in ways that build collective understanding of the phenomenon studied. (Cognition and Thechnology Group at Vaderbildt, 1990.) The design experiment is described in detail in Chapter 3.

2 Assessment of learning outcomes

Assessment in educational contexts can be seen as the conscious acquisition and interpretation of information about the knowledge and understanding, or abilities and

attitudes of another person. Assessment takes place in interaction, direct or indirect, with another. To some extent it is an attempt to know that person. Although assessment in instructional settings mainly takes place between teachers and students, one must not forget that students also assess teachers, other students, and that self-assessment is crucial for both students and teachers. (Rowntree 1987, 4).

Rowntree (1987) and Brown, Bull et al. (1997) identify various purposes of assessment. Assessment is used to select candidates for various kinds of educational opportunity and career; maintaining standards and quality-control; providing motivation and feedback to students and teachers. The term assessment is derived from *ad sedere* - to sit down beside. The implication of this etymology is that it is primarily concerned with providing guidance and feedback to the learner.

Rowntree (1987, 11) puts forward five dimensions of assessment which correspond to key activities in the process of assessment. The dimensions and questions underlying in these are as follows: 1) Decide why assessment is carried out; what effects or outcomes is it expected to produce. 2) Deciding what to assess. 3) Selecting truthful and fair methods. 4) Making sense of the obtained information; explaining, appreciating and attaching meaning to raw events of assessment, and finally 5) finding appropriate ways of expressing response and communicating it to the person concerned (and other people). In the present research, the assessment of learning outcomes serves mainly two purposes: it provides feedback and motivation to participants and it is used in the evaluation of the design experiment. The present assessment considers both learner's conceptual change and skills development in the two IR learning environments. The selection of methods for data generation and analysis are discussed in chapter 5 "Data and methods".

Brown, Bull et al. (1997, 17) discuss learning objectives and learning outcomes as guides to assessment tasks. Learning objectives are well-defined and tied to specific performance variables. Learning outcomes are more versatile and take into account prior learning and associated learning. The use of outcomes enables one to explore, in a more open fashion, what intended or unintended things have been learned. In the present research concept "learning outcomes" is used, because the intent is to analyze the whole spectrum of activities, not just comparing the instructional product to some pre-defined criteria. This approach is also in line with the conception of learning as a constructive process. In constructivism, assessment of learning outcomes is not an isolated process (like diagnostic, formative and summative evaluation) but it must be integrated into the learning process (see for example Jonassen (1992); Dochy and Moerkerke (1997)).

Several authors (see for example Biggs and Collis (1982); Reeves and Okey (1996); Dochy and Moerkerke (1997); Novak (1998); Biggs (1999)) have suggested new methods for the assessment of learning outcomes. These methods include authentic assessment, performance appraisal or assessment, learning logs or research diaries, and concept mapping. We have chosen concept mapping and performance assessment in the form of search session log as the methods of outcomes assessment in the present study. They offer the possibility to address both conceptual change and the development of searching skills. These assessment methods are described in detail in following sections.

2.1 Concept mapping

A concept map is a two-dimensional, hierarchical node-link diagram that depicts the structure of knowledge within a scientific discipline as viewed by a student, an instructor or

an expert in a field or sub-field. The map is composed of concept labels, each enclosed in a box or oval; a series of labeled linking lines, and an inclusive, general-to-specific organization. By reading the map from top to bottom, an instructor can:

- Gain insight into the way students view a topic;
- Examine the valid understandings and misconceptions students hold; and
- Assess the structural complexity of the relationships students depict.

Concept maps assess how well students see the "big picture". They have been used for over 25 years to provide a useful and visually appealing way of illustrating and assessing students' conceptual knowledge (Novak & Gowin 1984; Novak and Musonda 1991; Novak 1998).

2.2 Performance assessment

Three key features of performance assessment are: (1) students construct, rather than select, responses; (2) assessment formats allow teachers to observe student behavior on tasks reflecting real-world requirements; and (3) scoring reveals patterns in students' learning and thinking.

In performance assessment a learner is given realistic tasks and they are carefully observed while they carry out the tasks. Performance assessment tasks are popular for example in science education because they allow richer assessment that goes beyond testing the students ability to recall specific facts.

In information retrieval, search task completion includes the articulation of an information need into precise search keys and relationships supported by the IR system's query

language. Phases in these process as well as error types while searching can be categorized in various ways. According to Borgman (1986 ; 1996) a user must apply three kinds of knowledge and skills in IR process:

- Conceptual knowledge of the IR process - translating information need into a query.
This level also refers to understanding of query construction principles and for example selection of databases,
- Semantic knowledge of how to implement the query in the given system, and
- Technical skills i.e. syntactic knowledge in executing the query.

In human-computer interaction research, knowledge required and errors committed have been categorized into four levels (Foley & Van Dam, 1982; Shneiderman, 1992):

- *lexical level* includes knowledge of spelling and word forms,
- *syntactic level* covers the rules of command language, e.g. syntax,
- *semantic level* covers the meaning of commands and messages, and
- *conceptual level* includes the conceptual model of system and task.

Different levels of interaction have been identified also in applied search strategies. Fidel (1984, 1991) has identified two types of searchers based on their search strategies.

Conceptualists make conceptual moves while searching, e.g., search reformulation through search keys representing broader or narrower concepts. Operationalists rely on operational moves, e.g., applying time and language constraints and making use of record structure.

Categorizations of the knowledge required vary in terminology within various approaches. In the present performance assessment of IR skills these approaches are used in relation to search assignments and assessment criteria grounded on the data. In the development of IR

know-how, several factors can be identified, for example: the analysis of information need, selection of information sources, understanding of principles of information storage and retrieval techniques, evaluation of retrieved information etc. (Halttunen 2003a). In the present performance assessment it is suitable to focus on the factors present in a learning situation where search tasks as well as information sources are predefined. These factors are 1) translation of the search assignment into query, 2) implementation of query in given system, using appropriate system features and correct syntax in query execution, and 3) reflection and evaluation of queries and search results and possible modification of queries. Evaluation related to the conceptual level, i.e. participants' conceptual model of system and task, in performance assessment is ignored because it would be impossible to make such interpretations based on transaction logs.

3 A design experiment on IR learning environment

Understanding how technology and pedagogical solutions can best support student learning in diverse learning environments remains a crucial line of educational research and development. Finding a suitable approach to rapid technological change and the identification of best practices are core ideas of "design experiments". Collins (1992) describes an educational research experiment carried out in a complex learning context, which explores how a technological innovation affects student learning and educational practice (see also Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). The goals of design experiments should be design and implement innovative learning environments and simultaneously understand salient aspects of human cognition and learning involved with those innovations.

Design experiments:

- address learning programs involving important subject matter,
- are usually mediated by innovative technology,
- are embedded in everyday social contexts which are often classrooms, homes and workplaces where it is hard to control unanticipated events,
- account for multiple dependent variables
- do not attempt to hold variables constant, but the idea is to identify many variables and the nature and the extent of their effects, and
- evaluate different aspects of the design and develop a profile to characterize the design in practice.

The following paragraphs describe the context and instructional elements of our design experiment on IR learning environments.

The course "Introduction to Information Retrieval" (6 ECTS credits) in the Department of Information studies at University of Tampere is intended for first year undergraduate students. The course provides an overview of information storage and retrieval as practice and research area mediated with four instructional elements. First, lectures are given on basic concepts of information storage and retrieval. Themes like production of databases, matching, metadata, query formulation, and evaluation are covered. Second, weekly web-exercises concentrate on putting the themes covered in lectures in to practice. Every participant is working on these exercises making use of web-based tools and resources. Exercises are reported on web-forms, which are posted to the teacher. Third, tutored exercises in the classroom cover various aspects of information retrieval systems and their effective use. In these sessions, groups of 8-10 students work in pairs. Various kinds of search services are used (OPACs, union catalogues, article reference databases, full text

databases, Internet search engines and directories). Fourth, a course feedback web-form is filled out at the end of the course. This feedback covers three main areas: 1) course design and teaching methods, 2) the learner's self evaluation and role in the course and 3) the teachers role in the course. About 120 students attended this course and a lecturer and two tutors taking care of a part of the tutored sessions provided instruction. The outline of the course with different instructional elements and research activities is presented in Appendix 1.

In order to evaluate the instructional design and its experienced and observed effects on performance and learning a quasi-experiment was carried out. Tutored exercises were carried out in two different ways. In the *traditional learning environment*, different operational search systems (OPACs, union catalogues, article reference databases and full-text databases) were used to demonstrate the basic functions of IR systems. In the *experimental setting*, full-text newspaper articles from local newspaper along with a press image database were used. These sources were used through the IR Game, a system which offers feedback to the searcher on the effectiveness of queries based on recall-base. Various ideas of scaffolding and anchored instruction were applied in the experimental learning environment. Halttunen (2003b) provides detailed descriptions of these pedagogical solutions and instructional design summarized in Table 1.

Table 1. Summarization of differences in traditional and experimental learning environment.

	Traditional instruction	Experimental instruction
Instructional design	Searching databases	Searching in context. Anchored instruction.
	Unintentional scaffolding in the classroom.	Intentional scaffolding in the classroom and in the IR Game.
Performance feedback	Provided by tutor, no performance feedback in	Performance feedback in the IR Game and by tutors.

	operational systems used.	
Systems used	Six operational systems (85% of time) and the IR Game (15%).	The IR Game (85%) and two operational systems (15%)
Timing	Six one hour sessions.	Three two hour sessions.

Shadish, Cook and Campbell (2002) describe quasi-experiments to have participants' self-selection to treatment groups, a manipulable cause and enumeration of alternative causes. In the present research project, study participants were self-selected in to the groups, but their learning styles, prior conceptions of IR and educational/professional background were analyzed. Based on this analysis the groups were similar in all respects measured (see Halttunen, 2003a). Moreover, the students' self-selection in to the groups was based on each student's personal study schedule (which group fits best) and they were not aware of the nature of different groups. Therefore self-selection does not adversely affect our ability to make inferences from differences in learning observed between the groups. The manipulable cause in this study is the difference between the traditional and experimental learning environment as described in Table 1.

4 Research questions

Comparison of learning outcomes in the two IR learning environments implemented in the present design experiment form the basis for evaluation in the present project and contribute to our knowledge of IR teaching and learning. The specific research questions are:

- How students' conceptions of IR know-how developed during the instruction?
- What was the level of IR skills in the end of the instruction?
- Were there differences between groups studying in the traditional or the experimental learning environment in the development of conceptions or IR skills.

- Were there differences in the development of conceptions or IR skills related to student status (information studies major/minor), learning style, prior conceptions of IR know-how, or the academic discipline that they are studying as their major subject?

5 Data and methods

Multiple authentic datasets were gathered in the design experiment of IR learning environment to answer the research questions put forward above. These are: (1) a short essay describing students' conceptions of IR in the beginning of the course; (2) a questionnaire on conceptions; (3) a learning style inventory; (4) search logs produced in tutored exercises; (5) a short essay describing students' conceptions of IR in the end of the course; (6) empathy-based stories describing students' subjective learning experiences; and (7) course feedback. Background information on the participants' student status (studying IS as a major or a minor subject), domain of major subject (social sciences, humanities, science) were also gathered. Data gathering during the course is presented in Appendix 1. Multiple data sets allow data triangulation. There is the comprehensive collection of all seven datasets from 57 students of the total of 120 students participating in the course.

In the present article datasets 1, 4, and 5 are used to answer the research questions put forward above. These data sets support data triangulation to describe both conceptual change (essays) and skills development (log-files) in the two learning environments. In the same design experiment, data were also generated and analyzed on conceptions of IR know-how (1-2), learning styles (3), and learning experiences and performance (4, 6-7). These datasets were used in other studies of the present research project (see Halttunen 2003a; Halttunen 2003b).

The first body of data in the present article consists of essays (n=57) written in the very beginning of the course. These essays were first analyzed in order to investigate students' prior conceptions on IR know-how and their implication of the design of learning environment. Results of this study are reported by Halttunen (2003a). In the present research this data forms a baseline for the analysis of change of conceptions during instruction. The essays were re-analyzed for this purpose.

The second body of data consist of essays (n=57) written in the end of the course. These essays represent students' conceptions of IR know-how in the end of the instruction. They are used to analyze the development of conceptions i.e. conceptual change in comparison to the prior conceptions.

The third dataset consists of transaction logs of search sessions (n=24) of the Dialog-system used in the last session of tutored exercises. These 24 session logs represent searches by 57 students. The total number of queries was 5110 from which a sample was generated. The sample consists of 780 queries from the experimental groups (12 search sessions) and 743 from the traditional groups (12 search sessions). These files cover the same four exercises in both groups. The Library and Information Science Abstracts (LISA) database was used in this exercise. The session consisted of four search assignments (see Appendix 2), which were introduced to participants along the introduction and material of system features, which were appropriate for assignment. Forty-five minutes were spent finishing these exercises. The search session was the last of total of six hours of tutored search exercises within six weeks.

The first author who attended lectures and exercises of the course on a regular basis collected the data. He and the lecturer introduced data collection procedures and the aims of the research to the participants. Data collection was planned as an integral part of the course, not just as an artificial extra part. Writing an essay on prior conceptions of IR in the beginning of the course served as an orientation, and it created an interest in the topic at hand. Essays written in the end of the course were used to reflect and summarize students' conceptions of IR know-how.

Students wrote essays (45 minutes allocated) in the very beginning of the first lecture and in the one but last lecture as an in-class assignment. The timespan between essay writing sessions was eight weeks. The first round of data collection preceded any formal instruction in this course that might have influenced the findings. Instructions on essay writing in both cases were put forward as: "Write an essay-type of text, in which you present your own description of information retrieval know-how. You can approach the topic by identifying different kinds of skills, knowledge, elements etc. which, in your opinion, belong to IR know-how".

The essays were analyzed through concept mapping. Though concept-maps were introduced as vehicles for instruction, studying and assessment in the situations where they are used by teachers or students (Novak & Gowin 1984; Novak 1998), they can also be used as analytical methods for researcher in qualitative inquiry (Miles & Huberman, 1994; Åhlberg, 2001). In the present article learning outcomes in conceptual level were analyzed by constructing concept-maps of student essays. Concept-maps represent students' conceptions of IR know-how and were constructed by the first author based on students' own conceptions presented in their essays. Concepts referring to IR know-how were

extracted and labeled from the essays, as well as connections between concepts. The present lecturer of the course coded also half of the essays. The inter-rater consistency was 85.7 % regarding main concepts and 66,6 regarding sub-concepts at all levels. Reliability of coding is seen as good at the level of 80-85 % (Patton, 1990) Top-level concepts were identified within essays as new themes, which did not consist of examples or enriched definitions or descriptions of concepts presented earlier in the essay. Concepts were identified with qualitative coding of themes and concepts (Goffey & Atkinson, 1996, 26-51). Concepts were organized into hierarchies based on textual expressions in the essays. Figure 1 is an example of a concept-map.

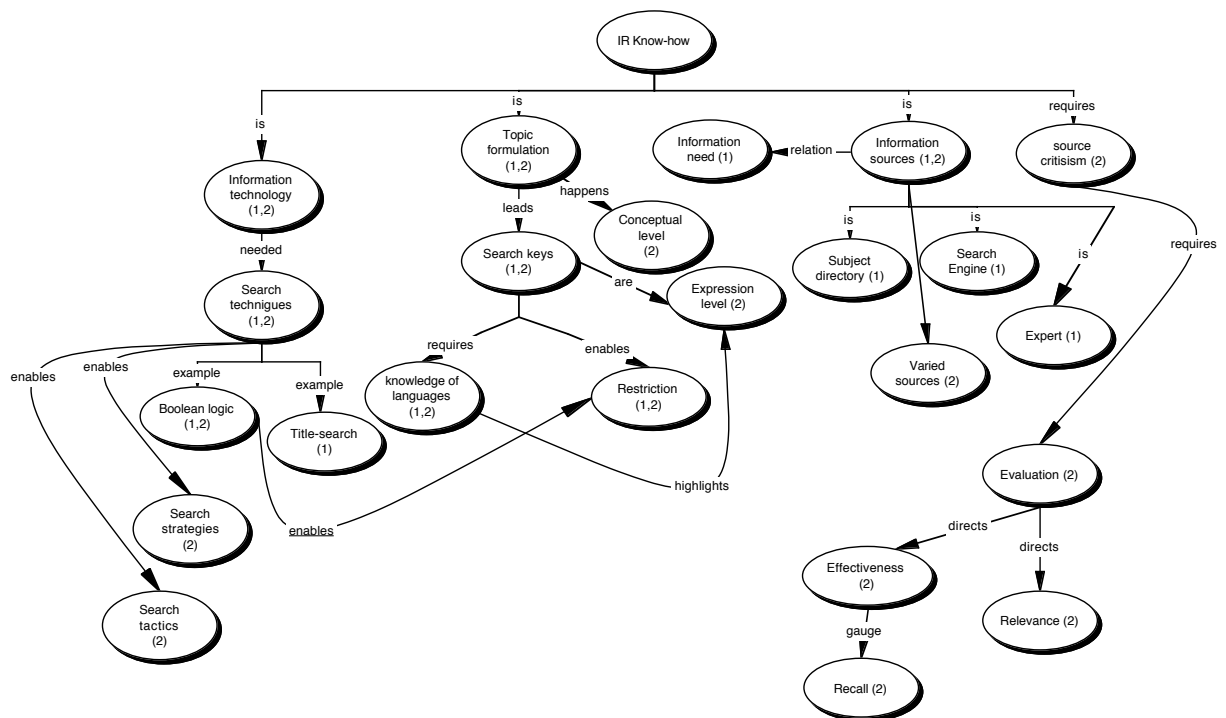


Figure 1. Concept-map created on essays of participant 05.

In Figure 1, text within ovals represents concept labels. Concepts presented in first essays are marked with (1). Concepts added in the second essay are marked with (2) and shading.

Concept labels with (1,2) are present in both essays. Links between subconcept hierarchies represent connections, effects, etc. between concepts. The assessment took account of the following (labels in the parenthesis refer to column labels in Tables 3–6.):

- The number of concepts in the beginning of instruction (beginning),
- The number of new concepts presented in second essay (new),
- The number of concepts in the end of instruction (end),
- The difference between the number of concepts in the beginning and end of instruction (difference),
- The number of concepts which remained same along the time (stable),
- The number of concepts that were ignored or changed fundamentally along the instruction (changed/ignored),
- The number of top level concepts in the end (top level)
- The number of new top level concepts presented in the second essay (new top level),
- The number of links between concept hierarchies in the end (links),
- Maximum depth of hierarchy levels in the end,
- The number of concepts per different levels of hierarchy in the end,
- The number of concepts per top-level concepts i.e. hierarchies in the end, and
- The level where new concepts were introduced.

Novak and Gowin (1984) have presented well-established scoring scheme to evaluate concept-maps. In this scheme 1 point is given for each correct relationship (i.e., concept-concept linkage); 5 points for each valid level of hierarchy; 10 points for each valid and significant cross-link; and 1 point for each example. In our approach we analyze different aspects of maps separately, because the researcher analyzes and draws maps based on student essay-like texts. We also want to give a more precise view of the change in

concepts, not just an overall grading. Our approach is in line with the ideas of how learners' conceptual models develop. Vosdianou (1994) has argued that mental models change in two ways, either by enrichment or revision of mental models. Enrichment refers to the addition of new information to existing knowledge structures while revision is needed when the information is inconsistent with existing beliefs or presuppositions. These two forms of learning identified in the cognitive studies of conceptual change derive from the Piagetian concepts of assimilation and accommodation. In our analysis the introduction of new top level concepts represents revision while adding concepts into existing hierarchies represents enrichment of knowledge structures.

In the last search session of the tutored exercises, log-files were collected by saving the search history from Dialog session into an HTML-file. Both students and tutors used search history throughout the session as a tool to analyze searching behavior.

Performance assessment is based on the analysis of log-files of search sessions at the tutored exercises. The Dialog retrieval system and the LISA-database was used in these sessions. Performance was assessed through analyzing problems and errors in queries and effectiveness of queries. A scoring scheme was devised for log-file analysis. The first author identified all possible errors which students had committed during each session. This raw data was categorized with the aid of previous research and literature of interaction problems in IR and OPAC systems and knowledge types in human-computer interaction (see for example Borgman, 1986; Borgman, 1996; Foley & Van Dam, 1982; Shneiderman, 1992; Sit, 1998). The present lecturer of the course also analyzed half of the log-files. The inter-rater consistency was 86,4 % regarding identification of errors and 89 % regarding

categorization. Because of the instructional assessment we excluded several factors of previous studies of operational environments, for example, database selection and information need formulation to search task, and concentrated on factors directly present in learning assignments. The analysis scheme contains following elements: 1) semantic and syntactic knowledge; 2) topical knowledge and functional (system) knowledge. Errors were categorized with the help of the matrix presented in Table 2.

Table 2. Matrix of error types.

	Topical knowledge	Functional (system) knowledge
Semantic knowledge	Missing facet Inappropriate search keys	Non-optimal fields Neglect phrase indexing
Syntactic knowledge	Misspelling and typos Problems in truncation	Misspelling of commands ; bad command sequences

6 Results and discussion

The following seven tables present the results of the analysis of students' conceptual change and the assessment of IR skills after studying in the IR learning environments for nine weeks (19.9.–21.11.2000). Tables 3–7 present conceptual change and the data in each table are enriched with qualitative description of concept maps. Performance assessment of searching skills is presented in Tables 8–9.

6.1 Analysis of the essays describing IR know-how

Major trends in students' conceptual understanding of IR know how developed, as natural in successful learning, in the following ways:

- New concepts were included in their conceptions,
- Old concepts were ignored or modified in learning process,
- Some concepts remained stable,
- New concept were added in various levels of hierarchy, and
- New linkages added between concepts sub-hierarchies represent advanced knowledge structures.

The following categories of concepts were presented in the essays. Knowledge of search process (information needs, information sources, information and document types, storage and retrieval systems and methods, evaluation of search results, access and use); meta-cognitive approaches (planning and management of search process); intermediaries and their qualities in IR activities; and finally other related factors for successful information searching (IT and linguistic know-how, general knowledge and all-round education).

Students also presented concepts related to connections of general concepts of information studies, e.g. definitions of information seeking and information retrieval, and some miscellaneous concepts.

Students presented an average of 11,0 concepts describing IR know-how in the beginning and 13,6 in the end of instruction – an increase of 2,6 concepts. Nearly eight (7,5) new concepts were added and 6,1 remained same. Almost five (4,9) concepts were modified fundamentally or ignored altogether. The average number of concepts presented altogether

was 18.5. Table 3 categorizes students by the number of concepts they expressed in the beginning and the end of instruction.

Table 3. Number of concepts presented in the essays

Concepts per essay	# students	# students
	beginning	end
1-5	7	1
6-10	18	18
11-15	23	18
16-20	8	14
21-25	1	6
Students total	57	57

The overall trends in concept development are described in the following. In prior conceptions of IR know-how the most frequent concepts were related to information sources, IR methods, information needs, and assessment of the information found. Concepts related to information sources and analysis of information needs were reduced in number along the instruction, and concepts related to planning and management of search process as well as linguistic skills were more emphasized in second essay. Also concepts describing IR methods and evaluation of search results were enriched along the course. Concepts related to computer skills and the characteristics of good information searchers remained quite stable in students' conceptions. The number of concepts related to general knowledge, intermediaries and information storage and organization decreased generally, although these concepts were more essential to some groups as we can see later in the results. The number of main concepts at the two phases of instruction is presented in Appendix 4.

Enrichment of concepts related to IR methods and management of search process is naturally in line with course goals.

6.2 IR learning environment and conceptual change

Comparison of students studying in the experimental and traditional learning environment in tutored exercises is presented in Table 4. The table also gives the total average number of concepts across all participants.

Table 4. Average number of concepts by groups based on groups and instructional design

	beginning	new	stable	end	differenc e	changed/ ignored	top level concepts	new top level	links
experimental group	12,0	9,3	6,0	15,3	3,3	6,0	5,8	1,9	2,5
traditional group	10,6	6,8	6,2	13,0	2,4	4,4	6,2	1,9	1,2
Total average	11,0	7,5	6,1	13,6	2,6	4,9	6,1	1,9	1,6

The results indicate that students in the experimental group presented a somewhat larger number of concepts already in the beginning of instruction, but they also introduce much more new concepts as well reach a more exhaustive view of IR know-how in the end. The experimental group changed fundamentally or ignored more concepts while studying than students in the traditional learning environment. Students in the traditional group were somewhat more stable in their conceptions throughout the instruction. A remarkable difference is the number of linkages between concept hierarchies.

Qualitative analysis of conceptual change in these groups revealed the following trends.

The role of information sources and computer skills diminished more in experimental

group. The experimental group paid more emphasis on linguistics as element of IR know-how than students in traditional group. The same trend was present also regarding the role of intermediaries as well as planning and management of searching process. In other aspects the trends were similar or conceptions remained stable.

6.3 Prior conceptions and conceptual change

Students' prior conceptions of IR know-how were analyzed in an earlier article in the current research project (see Halttunen, 2003a). One result of the study was the identification of five qualitatively different ways to approach IR. These conceptions emphasized different phases of search process, namely:

- *process identifiers* covered phases of search process most exhaustively,
- *source identifiers* concentrated on identification of information sources,
- *searchers* paid emphasis to search techniques and tools,
- *problem formulators* concentrated on the analysis of search tasks, and
- *assessors* emphasized evaluation of search results and information found.

Figure 2 and Appendix 3 present the change of conceptions of these groups. It can be seen, that groups divided their conceptions more equally to various phases of search process at the end of the course. The most radical changes can be identified in groups of "problem formulators" and "assessors". All groups broadened their views to a better balanced one. In their prior conceptions the groups were significantly different ($\chi^2=52,848$; $p=0,0001$), but became more similar to each other in the end of instruction ($\chi^2=8,191$; $p=0,9905$). Test statistic based on Appendix 3.

Type	Phase	Info	Info	IR Met	Info	Assess	Access
------	-------	------	------	--------	------	--------	--------

	Need	Sources	hods	Storage	ment	and use
Process identifiers						
-end conception						
Source identifiers						
-end conception						
Searchers						
-end conception						
Problem formulators						
-end conception						
Assessors						
-end conception						

Legend: shading expresses the share of students expressing each concept category as follows: 100%=black ; 99-67%=dark grey ; 66-34%=medium grey ; 33-1%= light grey ; 0%=white

Figure 2. Phases of search process identified by groups based on their process views.

Analysis of conceptual change within these groups is presented in Table 5.

Table 5. Conceptual change by groups based on prior conceptions

Prior conception of IR know-how	Average number of concepts								
	beginning	new	stable	end	difference	changed/ ignored	top level concepts	new top level	links
Process identifiers	10,9	6,7	7,5	14,2	3,3	3,5	5,6	1,5	1,1
Source identifiers	9,7	8,8	5,4	14,3	4,6	4,3	6,4	2,6	1,8
Searchers	11,4	8,2	6,5	14,7	3,3	4,9	6,3	1,7	1,7
Problem formulators	14,3	6,0	6,3	12,3	-2,0	8,0	5,5	1,3	1,7
Assessors	10,6	6,0	5,0	11,0	0,4	5,6	6,0	1,8	1,6

Students who covered well phases of search process in their prior conceptions - *Process identifiers* - did not radically change their conceptions. They both changed/ignored least concepts and enriched their conceptions related to the end phases of the search process. *Source identifiers* presented least concepts in their essays in the beginning of the course, but they introduced the greatest number of new concepts and also new top-level concepts in their essays in the end of semester. They were also students who had largest difference between number of concepts in the beginning and end of semester. *Source identifiers* both enriched and revised their conceptual understanding of IR. They introduced concepts related to IR methods, information needs and assessment, while they ignored concepts related to access and use of information.

Searchers presented the second largest number of concepts in the beginning concentrating on IR methods and assessment of search results. They introduced, however, some new top-level concepts related to information needs, storage, and resources. They consolidated their overall view of search process. Interestingly, *problem formulators*, i.e. participants who paid emphasis on the beginning of search process (analysis of information needs, topic selection and problem formulation) were those who changed their conceptions of IR know-how most radically during instruction. During the semester their conceptions of IR changed most comprehensively and introduced concepts that were not present in their prior conceptions. In the end they presented fewer concepts than in the beginning - they raised the abstraction level of their concepts. *Assessors*, who emphasized the assessment of information content and relevance in their prior conceptions also changed their concepts, but the number of concepts remained at the same level as in the beginning although they introduced the second largest number of new top level concepts. *Assessors* presented especially a variety of concepts related to IR know-how, which cannot be categorized as

phases in the search process. These are for example, computer and linguistic skills, role of intermediaries, and personal qualities of information seekers.

There is some qualitative differences between concepts related to prior conceptions in traditional and experimental learning environments. First, *problem formulators* in experimental group covered phases of search process more exhaustively than students in traditional group. They paid emphasis especially issues related to information storage and assessment. Second, both *process identifiers* and *source identifiers* enriched their concepts related to assessment more in experimental learning environment.

6.4 Learning styles and conceptual change

In the previous study (Halttunen, 2003a) concerning prior conceptions of IR also students' learning styles were analyzed with Kolb's (1976, 1984) learning style inventory (LSI). The distribution of participants was the following: students leaning toward concrete experience (10), reflective observation (26), abstract conceptualization (16), and active experimentation (5). (Halttunen, 2003a). The distribution of the number of concepts and links between hierarchies classified by learning style is presented in Table 6.

Table 6. Conceptual change classified by learning style

Learning style	Average number of concepts									
	beginning	new	stable	end	diffe- rence	changed/ ignored	top level concepts	new top level	links	
Concrete experience	9,5	8,8	4,4	13,2	3,7	5,1	6,9	2,4	1,2	
Reflective observation	11,5	7,9	7,0	14,9	3,5	4,4	5,8	1,9	1,9	
Abstract conceptualization	10,9	6,3	5,9	12,1	1,2	5,1	5,9	1,5	1,3	
Active experimentation	11,8	7,0	5,6	12,6	0,8	6,2	6,2	2,0	1,8	

Participants with learning style and mode that leans toward "concrete experience" introduced more new concepts and more new top level concepts than students with other learning styles. They also presented more top-level concepts altogether than the others. Participants with "reflective observation" learning style kept more concepts stable and changed or ignored fewer concepts than participants in the other learning styles. Participants with "abstract conceptualization" introduced fewer new concepts and top-level concepts than other groups. Students with learning style "active experimentation" seem also active in changing and ignoring concepts. They introduced new concepts related to linguistic talent, publishing and the processes of planning and executing search tasks. Concepts related to access and use of information were ignored by the groups "abstract conceptualization" and "active experimentation". There were some trends related to learning styles, students belonging to learning styles "concrete experience" and "reflective observation" introduced new concepts related to information needs and sources, while slight opposite trend is characteristic for other learning styles.

There were some trends in students' conceptual change regarding the learning environment they were studying. The overall trend related to all learning styles was the emphasis on linguistic aspects of IR in the experimental learning environment. Also assessment of information was emphasized in experimental environment in other learning styles than concrete experience. Students with learning styles reflective observation and abstract conceptualization paid more emphasis on information sources in experimental learning environment than students with other learning styles or environment.

6.5 Student status and conceptual change

There were no great differences in the number of concepts between students studying Information Studies as their major or minor subject (Table 7). However, the domain of the major subject and through that the background and orientation to studies, revealed interesting distinctions.

Table 7. Average number of concepts by groups based on student status and domain of study.

Student status, domain of study	Average number of concepts									
	beginning	new	stable	end	diffe- rence	change	top level concepts	new top level	links	
IS as major subject	11,0	7,7	6,0	13,7	2,7	5,0	6,0	1,8	1,7	
IS as minor subject	11,0	7,3	6,2	13,6	2,6	4,8	6,2	2,0	1,5	
Social Sciences	11,0	7,5	6,1	13,6	2,5	4,9	5,9	1,8	1,7	
Humanities	9,6	8,9	6,1	15,0	5,4	3,5	6,4	2,4	1,4	
Sciences	12,0	6,2	8,8	15,0	3,0	3,2	5,8	1,5	2,2	
Master's program	12,3	6,6	4,1	10,8	-1,5	8,1	6,5	1,5	1,0	

Social science students, which were mainly studying information studies as major subject, differ from other groups in reducing the number of concepts related to information needs, IR methods and computer skills, while the other groups presented more concepts related to these topics. Major students emphasized information needs and IR methods in the beginning possibly due the entrance examination to information studies, where they had to master an exam book partially covering the course.

Humanities students had the fewest concepts in the beginning, but they introduced many new concepts at all levels of concept map. New top level concepts were introduced more

frequently than in any other group. These students introduced more concepts related to IR methods and information assessment than the other students, and reduced radically their focus on publication types as part of IR know-how. Humanities' students did not regard computer skills as important as science students, though there was a slight increase in the number of these concepts.

Science students started with quite thick descriptions, although they missed concepts related to linguistic and intermediary aspects of IR know-how. They did not introduce as many new concepts as the other students did. Their total change was the smallest, but in the same time they introduced more linkages between hierarchies than other groups. The major changes in their concepts are related to increased focus on information needs and sources, computer skills, linguistic aspects and diminishing focus on assessment, access and use.

Participants' from *Master's program* of networked information services differ clearly from other groups. Their background is either journalism, computer science or information studies and they have quite a long work-history and have finished their basic studies at least some years ago. They presented the largest number of concepts in the beginning, but radically changed or ignored presented concepts during instruction. They did not present many new top level concepts, but modified especially concepts related to information needs, sources and methods, and less so to concepts related to computer and linguistic skills, intermediary functions and management of search process. They clearly raised the abstraction level of their concepts.

Qualitative differences in concept according to student status and learning environments were following. First, major students emphasized IR methods both in traditional and

experimental learning environment. Second, students studying in experimental groups provided richer description on evaluation, linguistic aspects and information use as elements of IR know-how than students in traditional groups. Third, traditional groups emphasized more general knowledge and information storage as important elements of IR know-how. Analysis based on domain of study related to learning environment revealed one interesting feature. Although students in experimental learning environment tended to emphasize linguistic aspects of IR, the computer science students in experimental learning environment did not find this aspect important at all.

6. 5 Performance assessment of search sessions

The previous section presented an analysis and assessment on the development of students conceptual change expressed in their short essays written in the beginning and end of instruction. The following tables present the results of the analysis of log files which describe student performance in four search exercises described in "data and methods" section. First, we present the results of performance assessment of search sessions based on the analysis of encountered problems and errors. Second, we present an overall grading of learning outcomes in performance assessment.

Table 8. Distribution of errors by error type and instructional group (n=228)

	Traditional group		Experimental group	
	#	% of total	#	% of total
Semantic knowledge errors	84	36,8	56	24,6
<i>Topical knowledge</i>	57		20	
missing facet	13		2	
combination of facets, keys	15		10	
inappropriate search keys	29		8	
<i>Functional knowledge</i>	27		36	

neglect phrase indexing	27		36	
Syntactic knowledge errors	59	25,9	29	12,7
<i>Topical knowledge</i>	52		26	
misspelling, typos	4		0	
truncation	48		26	
<i>Functional knowledge</i>	7		3	
bad commands	7		3	
total	143	62,7	85	37,3

Performance assessment revealed that both the traditional and the experimental groups made semantic and syntactic errors. Though error rate varies much in different categories and differences are statistically significant ($\chi^2=15,586$; $p=0,0014$). Statistics are based on number of errors in topical and functional knowledge both in semantic and syntactic knowledge. (Italics in the Table 8.). The traditional group made more semantic knowledge errors than participants who studied in the experimental learning environment. These errors were related to the process of transforming a search assignment into a query. A missing facet indicates insufficient analysis of the search assignment and insufficient identification of central concepts and facets in the request. Inappropriate search keys refer to situation where a searcher does not take into account the discourse of the database they are using. Errors in combining search keys either with Boolean or proximity operators are regarded as semantic errors, because inappropriate use of these influences the meaning of the search assignment. These errors could also be categorized as syntactic errors, but based on data they are more related to the analysis of search assignments than difficulties of understanding of search syntax of the system. Students in experimental group had become used to analyze search assignments more deeply, probably because of more general anchored instructional approach. They also had become used to put emphasis on the identification and selection of search keys because of the wide variety of expressions in

full-text newspaper database they used in the exercises. Also the feedback and scaffolding of IR Game probably made them more aware of the effects of erroneous combinations of search keys than the students using operational systems in the traditional group.

Students from both learning environments made quite the same number of syntactic knowledge errors. Participants in both groups failed to use the properties of the author field in first search assignment, but after the identification of possible variation of name forms and phrase indexing, they managed to overcome this problem. Lexical errors were present in traditional group. It seems that anchored and scaffolded instruction had more effect on the overall management of search process and analysis of search tasks than was able to produce transferable syntactic knowledge.

We assessed the learning outcomes of the basics of IR with performance assessment as a whole by grading performance in search assignments on a scale from 1-4. Grading is based on a scheme where in every assignment the best result set was taken into account. Students were able to receive 1 point when they reached > 70 % and 0,5 point 50–70 % of result sets formed by lecturer as examples of various approaches to search assignments. Less than 50 % yielded 0 points. Because information search tasks seldom have one ultimate solution, we used a collection of teacher's example queries and results as a frame of reference in the evaluation and compared students' strategies to similar examples, which are presented in Appendix 2. In this phase we did not consider the number of iterations to reach the best results, nor was the grading affected by errors made in some queries. The overall grading of performance assessment by groups is presented in Table 9.

Table 9. Assessment of search success in search sessions (scale 1=poor, 4=excellent)

Grading of search assignment	traditional	experimental
	group	group
	# students N=12	# students N=12
1	0	0
1,5	2	0
2	4	2
2,5	1	3
3	3	1
3,5	1	5
4	1	1
average	2,5	3

Students studying in the experimental learning environment received better grading by 0,5-points than the participants in the traditional environment. Notably six of them received excellent grading ($\geq 3,5$) while the same grading was only achieved by two students in the traditional group. Opposite trend can be found also in the beginning of the scale. Although, difference between groups is not statistically significant.

7 Conclusions

The contribution of this article is the assessment of learning outcomes in two information retrieval learning environments. In the experimental environment, instructional approaches like anchored instruction and scaffolding were utilized. Scaffolding was provided by an educational software tool, the IR Game, and by tutors in the classroom setting. Anchored

instruction was based on the idea of IR activities taking place in journalistic work-task situation. The assessment of learning outcomes took place at two levels. Student's conceptual change was analyzed through concept mapping based on student essays. The development of IR skills was investigated with transaction logs representing user-system interaction while completing a search assignment.

The experimental group changed fundamentally, or ignored, more concepts while studying than students in the traditional learning environment. The traditional group was somewhat more stable in their conceptions throughout the instruction. The emphasis of information sources and computer skills diminished more in experimental group. The experimental group paid more emphasis on linguistics as an element of IR know-how than students did in the traditional group. The same trend was also related to the role of intermediaries as well as planning and management of searching process. Instructional approaches, i.e. anchoring and scaffolding, applied seem to be promising strategies to stress the importance of planning and management of search process, as well as putting emphasis on important linguistic aspects of IR.

The analysis of conceptual change that relates to students' prior conceptions revealed successful learning outcomes. Regardless of sparse conceptions in the beginning of the instruction, the participants were able to form an overall picture of IR activities.

Conceptions of *process identifiers*, *source identifiers*, *searchers*, *problem formulators* and *assessors* changed to more consistent with each other. The instructors of IR could make these prior conceptions visible to learners and use them as tools to construct conceptual understanding of various aspects of IR.

There is some qualitative differences of students' conceptions in traditional and experimental learning environments. First, *problem formulators* in experimental group covered phases of search process more exhaustively than students in traditional group. Second, both *process identifiers* and *source identifiers* enriched their concepts related to assessment more in experimental learning environment.

The learning styles analyzed in the study seem to have some effect on students' conceptual change, whereas there were no great differences in the number or qualities of concepts between students studying Information Studies as their major or minor subject. However, the domain of the major subject created interesting differences. Humanities' students introduced new top level concepts more frequently than any other group. Science students were most stable and their total change was the smallest, but in the same time they introduced more linkages between hierarchies than other groups. Students of Masters Program of Networked Information Services radically diminished the number of concepts during the instruction and raised the abstraction level of their concepts. The results indicate that different student groups implement different strategies to form a usable conceptual framework for further studies. Making use of prior conceptions and the identification of formation strategies could be used as successful instructional approach in IR.

In the experimental learning environment the overall trend related to all learning styles was the emphasis on linguistic aspects of IR. Also assessment of information was emphasized in experimental environment in all other learning styles than concrete experience.

Qualitative differences in concept according to student status and learning environments were following. First, major students emphasized IR methods both in traditional and experimental learning environment. Second, students studying in experimental groups

provided richer description on evaluation, linguistic aspects and information use as elements of IR know-how than students in traditional groups. Third, traditional groups emphasized more general knowledge and information storage as important elements of IR know-how.

The development of IR skills was evaluated through performance assessment, which took place in the last session of tutored exercises. The IR system and database used in this session was new to all of the participants. There was a statistically significant difference in the error types which students encountered in these exercises. The traditional group made much more semantic knowledge errors than participants who studied in the experimental learning environment. These errors were related to the process of transforming a search assignment into a query. Students from both learning environments made quite the same number of syntactic knowledge errors. It seems that both groups were able to overcome problems with syntactic errors with active exploration, but semantic problems effected their overall performance since students in traditional environment were not able to reach as good search results as participants in experimental group.

Further research is needed to evaluate different kinds of approaches to IR learning environments and their design. First, because the benefits of anchoring and scaffolding are not categorical because the range of intervening variables and the difficulty of setting up a design experiment which tried to be naturalistic but at the same time tried to focus on a specific aspect. Second, there are very few studies related to instructional methods or assessment of learning outcomes of IR instruction although instruction is provided by, for example, online vendors, libraries, universities, and schools.

Appendix 1

Events, topics and research activities in the introductory course of information retrieval.

Week	Event	Topic	Research activities, data sets
1	lecture, web exercises	Introduction to the course theme and activities <ul style="list-style-type: none"> • IR as practice and research area • Short introduction to databases 	Essay (1) and questionnaire (2) on prior conceptions of IR know-how
2	lecture, web exercises	Information storage and retrieval I <ul style="list-style-type: none"> • Production of databases • Structure of databases, query languages 	Learning style inventory (3)
3	lecture, web exercises	Information storage and retrieval II <ul style="list-style-type: none"> • Query languages (cont.) • Search keys, search statements • Matching (best match, exact match) • How to research use of databases 	

	tutored exercises	Traditional groups <ul style="list-style-type: none"> • National catalogue of university libraries (LINDA) • Library of Congress (LOC) 	Experimental groups <ul style="list-style-type: none"> • Full-text database of local newspaper • IR Game 	Log-files of IR Game (4)
4	lecture, web exercises	Information organization I <ul style="list-style-type: none"> • descriptive cataloging • standards and formats (FINMARC) • future of cataloging 		
	tutored exercises	<ul style="list-style-type: none"> • EBSCOhost • full-text databases • browsing 	<ul style="list-style-type: none"> • no session 	
5	lecture, web exercises	Information organization II <ul style="list-style-type: none"> • subject description (keywords, descriptors, abstracting, indexing, classification) • Three levels of IR (concepts, expressions, strings) 		

	tutored exercises	<ul style="list-style-type: none"> • Article reference database (ALEKSI) • Use of thesauri 	<ul style="list-style-type: none"> • Full-text database of local newspaper • IR Game • Article reference database (ALEKSI) • Use of thesauri 	Log-files of IR Game (4)
6	lecture, web exercises	Search strategies and tactics Evaluation of ISR		
	tutored exercises	<ul style="list-style-type: none"> • IR Game • Relevance • Precision, recall 	<ul style="list-style-type: none"> • No session 	Log-files of IR Game (4)
7	lecture, web exercises	IR in the web environment		
		<ul style="list-style-type: none"> • Image retrieval • Web searching 	<ul style="list-style-type: none"> • No session 	
8	lecture, web exercises	Information organization of		
		<ul style="list-style-type: none"> • music • images 		

		<ul style="list-style-type: none"> • Dialog -system • LISA-database • Performance assessment 	<ul style="list-style-type: none"> • IR Game • Relevance • Precision, recall • Image retrieval • Dialog -system • LISA-database • Performance assessment 	<p>Log-files of IR Game (4)</p> <p>Log-files of LISA (4)</p>
9	lecture, web exercises	<p>Metadata and identification of electronic documents (e.g. Dublin Core, DOI)</p> <p>Parallel searching (Z39.50)</p>		Essay (5) on conceptions of IR know-how
10	lecture, web exercises	<p>Role of intermediaries in IR</p> <p>Future trends of IR</p>		Empathy-based story (6) of learning experiences
11	web exercises			Course feedback questionnaire (7)

Appendix 2 Search assignments in the tutored exercises using LISA database in the Dialog system.

Exercise	Assignment (translated from finnish)	Examples of good queries	# items found
1	What Carol Tenopir and William Hersch have written about evaluation of databases	ss (au=tenopir, c? OR au=hersh, w?) and (data(w)base? or database?) and evaluation?	11
2	Find material on evaluation of internet resources for your seminar paper. Evaluate the search results.	ss (internet? or world(w)wide(w)web? or www) and resource? and evaluation? and la=english	135
		ss (web(w)page?/de OR web(w)site?/de OR internet?/de or world(w)wide(w)web/de) and resource?(F)evaluation? AND la=english	80
3	Find articles concerning IR interfaces for children and young people.	ss (child? OR young(w)people OR youth?) AND interface? AND (information(w)retrieval OR "information storage and retrieval" OR search?)	24

		ss (child? OR young(w)people OR youth?) AND interface? AND (information(w)retrieval OR "information storage and retrieval")	10
4	Information on education and research on records management is needed. Material should be published after 1995. Try to reach good precision.	ss records(w)management? and (education? or train? or research?) and py>1995	147
		ss records(w)management?/de and (education?/de or train?/de or research?/de) and py>1995	54

Appendix 3. Number of students' conceptions in each phase of search process

Type	Phase	Info. need		Info. source		I R methods		Storage		Asses- sment		Access an use	
		#	%	#	%	#	%	#	%	#	%	#	%
		Proces identifiers		9	100	9	100	9	100	1	11	3	33
- end conceptions		8	89	9	100	9	100	4	44	4	44	3	33
Source identifiers		4	36	11	100	0	0	2	18	3	27	3	27
- end conceptions		5	45	11	100	11	100	2	18	4	36	1	9
Searchers		7	33	7	33	21	100	6	29	10	48	4	19
- end conceptions		15	71	17	81	21	100	13	62	12	57	3	14

Problem formulators	6	100	0	0	0	0	2	33	3	50	1	17
- end conceptions	6	100	6	100	6	100	3	50	3	50	1	17
Assessors	1	14	4	57	0	0	1	14	7	100	3	43
- end conceptions	4	57	7	100	7	100	4	57	6	86	2	29

Appendix 4. Main categories of concepts presented in the essays (n=57).

Concept	beginnin g	new	both	end	change in #
Information sources	38	7	28	35	-3
IR methods	32	29	25	54	22
Information needs	27	18	13	31	4
Evaluation	25	21	12	33	8
Computer skills	15	12	9	21	6
Intermediaries	9	8	1	9	0
General knowledge	8	10	2	12	4
General concepts of IS	7	2	1	3	-4

IR process	6	19	2	21	15
Qualities of information seeker	6	8	1	9	3
Information storage	6	7	1	8	2
Publication types and formats	6	5	3	8	2
Linguistic talent	4	15	2	17	13
Use of information	3	2	0	2	-1
Misc.	3	1	0	1	-2

References

- Biggs, J. (1999). *Teaching for quality learning at university: What the student does*. Buckingham: Society for Research into Higher Education & Open University Press.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning*. New York: Academic Press.
- Borgman, C. L. (1986). Why are online catalogs hard to use? : Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37(6), 387-400.
- Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.
- Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2001). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria Digital Earth Proto Type (ADEPT). *Library Trends*, 49(2), 228-250.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.
- Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher* 19(3), 2-10. Collins, A. (1992).

Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* . (pp. 15-22).Berlin: Springer.

Dochy, F., & Moerkerke, G. (1997). Assessment as a major influence on learning and instruction. *International Journal of Educational Research*, 27(5), 415-432.

Fidel, R. (1984). Online searching styles: a case-study-based model of searching behavior. *Journal of the American Society for Information Science*, 35(4), 211-221.

Fidel, R. (1991). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science*, 42(7), 515-527.

Foley, J. D., & Van Dam, A. (1982). *Fundamentals of computer graphics*. Reading, Mass.: Addison-Wesley.

Goffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: complementary research strategies*. Thousand Oakes: Sage.

Gutzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, 4(1), 1-44.

Halttunen, K. (2003a). Students conceptions of information retrieval: implications for design of learning environments. *Library and Information Science Research*, 25(3), 307-332.

Halttunen, K. (2003b). Scaffolding Performance in IR Instruction Exploring Learning Experiences and Performance in Two Learning Environments. *Journal of Information Science*, 29(5), 375-390.

Halttunen, K., & Sormunen, E. (2000). Learning information retrieval through an educational game: is gaming sufficient for learning? *Education for Information*, 18(4), 289-311.

Jacobson, F. F., & Ignacio, E. N. (1997). Teaching reflection : information seeking and evaluation in a digital library environment. *Library Trends*, 45(4), 771-802.

- Johnson, A. M. (2001). Library instruction and information literacy - 2000. *Reference services review*, 29(4), 338-361.
- Johnson, A. M. (2002). Library instruction and information literacy - 2001. *Reference services review*, 30(4), 359-389.
- Jonassen, D. H. (1992). Evaluating constructivist learning. In T. Duffy & D. Jonassen (Eds.), *Constructivism and the technology of instruction: a conversation*. (pp. 137-148). Hillsdale, NJ: Lawrence Erlbaum.
- Kolb, D. (1976). *The learning style inventory: technical manual*. Boston: McBer and Company.
- Kolb, D. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ.: Prentice-Hall.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook* (2 ed.). Thousand Oaks: Sage.
- Novak, J., & Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American educational research journal*, 28(1), 117-135.
- Novak, J. D. (1998). *Learning, creating, and using knowledge : concept maps as facilitative tools in schools and corporations*. Mahwah, NJ.: Lawrence Erlbaum Associates.
- Novak, J. D., & Gowin, B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park: Sage.
- Pausch, L. M., & Popp, M. P. (1997). *Assessment of information literacy: lessons from the higher education assessment movement*. Association of College and Research Libraries.[Available: <http://www.ala.org/acrl/paperhtm/d30.html>].
- Rader, H. B. (2000). Library instruction and information literacy - 1999. *Reference services review*, 28(4), 378-399.

- Rader, H. B. (2002). Information literacy 1973-2002: A selected literature review. *Library Trends*, 51(2), 242-259.
- Reeves, T. C., & Okey, J. R. (1996). Alternative assessment for constructivist learning environments. In B. G. Wilson (Ed.), *Constructivist learning environments: case studies in instructional design*. (pp. 191-202). Englewood Cliffs, NJ: Educational technology publications.
- Rowntree, D. (1987). *Assessing students: How shall we know them?* London: Kogan Page.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction* (2. ed.). Reading, Mass.: Addison-Wesley.
- Sit, R. A. (1998). Online library catalog search performance by older adult users. *Library and Information Science Research*, 20(2), 115-131.
- Sormunen, E., Halttunen, K., & Keskustalo, H. (2002). *Query Performance Analyser : a web-based tool for information retrieval research and instruction*. (Research Notes 1/2002). Tampere: Department of Information Studies, University of Tampere.
- Sormunen, E., Laaksonen, J., Keskustalo, H., Kekäläinen, J., Kemppainen, H., Laitinen, H., Pirkola, A., & Järvelin, K. (1998). *IR Game : a tool for rapid query analysis in cross-language IR experiments*. Paper presented at the Joint workshop on cross language issues in artificial intelligence & issues of cross cultural communication, Singapore.
- Vosdianou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4(1), 45-69.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of child psychology and psychiatry*, 17(2), 89-100.

Åhlberg, M. (2001). Käsitekartat tutkimusmenetelmänä [Concept-maps as research method]. In J. Aaltola & R. Valli (Eds.), *Ikkunoita tutkimusmetodeihin 1 [Windows into research methods 1]*. (pp. 59-68). Jyväskylä: PS-Kustannus.