

## Research Article

# Cognitive Load in eCommerce Applications—Measurement and Effects on User Satisfaction

Peter Schmutz,<sup>1</sup> Silvia Heinz,<sup>1</sup> Yolanda Métraiiler,<sup>2</sup> and Klaus Opwis<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland

<sup>2</sup> IV-Stelle Basel-Stadt, Lange Gasse 7, 4052 Basel, Switzerland

Correspondence should be addressed to Peter Schmutz, peter.schmutz@unibas.ch

Received 15 October 2008; Accepted 24 February 2009

Recommended by Richard Kline

Guidelines for designing usable interfaces recommend reducing short term memory load. Cognitive load, that is, working memory demands during problem solving, reasoning, or thinking, may affect users' general satisfaction and performance when completing complex tasks. Whereas in design guidelines numerous ways of reducing cognitive load in interactive systems are described, not many attempts have been made to measure cognitive load in Web applications, and few techniques exist. In this study participants' cognitive load was measured while they were engaged in searching for several products in four different online book stores. NASA-TLX and dual-task methodology were used to measure subjective and objective mental workload. The dual-task methodology involved searching for books as the primary task and a visual monitoring task as the secondary task. NASA-TLX scores differed significantly among the shops. Secondary task reaction times showed no significant differences between the four shops. Strong correlations between NASA-TLX, primary task completion time, and general satisfaction suggest that NASA-TLX can be used as a valuable additional measure of efficiency. Furthermore, strong correlations were found between browse/search preference and NASA-TLX as well as between search/browse preference and user satisfaction. Thus we suggest browse/search preference as a promising heuristic assessment method of cognitive load.

Copyright © 2009 Peter Schmutz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Within the past few years, the Internet has grown and shifted from an information medium to a workspace where users manage tasks of growing complexity. Information search, participation in online communities, multimedia sharing, selling, and buying are only a few of a wide range of online activities that users may perform. Taking full advantage of these possibilities places high demands on users and designers in equal measure. The users navigate through the Web, search for and retrieve information, have to prioritize and constantly make selections and decisions. The common task of buying a product in an online store may serve as a good example for the complexity of online activities; given that users have the goal of buying a book, they will visit an online bookstore, then navigate and search for the desired object by using either the navigation or the search facility. They will roughly remember the name and the appearance of its cover and will match this expectation against a set of distracting unwanted books. Having perceived the target

book, users will try to figure out how to put the item into the shopping cart and search for it. The checkout process finally involves another set of tasks such as reviewing items in the cart, retrieving the address from the long-term memory, and copying credit card information from the credit card, filling out forms and correcting errors reported by the shopping engine.

Performance and success in these tasks depend not only on the users' abilities, for example, knowledge and working memory capacity (e.g., [1]), but also on the task itself and the way the respective websites and tools are organized, that is, the visual layout of the page, the usability of the user interface, and its interaction design. In other words, the ease of completing a task is a function of task complexity and capacity of the users' cognitive resources. Thus, Web designers are faced with the highly demanding task of adapting these functionalities to the needs and characteristics of a growing and heterogeneous base of users.

Several Web design and usability guidelines have been elaborated (e.g., [2–4]; for an overview, see [5]) in order

to support the designers in this task. Many of them point to the need for taking into account the capacity limitation of the users and suggest reducing cognitive workload. Shneiderman's rules for design [6], for example, include, among seven other rules, the requirement for the reduction of short-term memory load. Mandel [7] points out the relevance of reducing users' memory load and provides nine principles for reducing cognitive load based on the knowledge of user behavior and cognition.

At the same time, different cognitive principles with potentially misleading interpretations such as the magical number seven [8], a limited processing capacity principle, are adapted in design guidelines. However, research concerning the effectiveness of these cognitive load reduction guidelines is sparse.

One of the major problems in exploring cognitive load is the problem of measurement. How can we assess cognitive load in a particular task and how can we test whether a principle may effectively reduce cognitive load? How can we measure the improvements in user efficiency or even user satisfaction? To none of these questions have clear answers yet been found by human-computer research.

In this paper we will therefore try to shed light on these questions by adapting fruitful aspects from instructional design and learning research, namely, in the Cognitive Load Theory (CLT) by Sweller [9]. CLT and cognitive load measurement will be covered in the theoretical background section.

## 2. Theoretical Background

### 2.1. Existing Research

*2.1.1. Cognitive Load Theory (CLT).* Based on works about problem solving (e.g., [10]), expert versus novice research (e.g., [11]), and learning (e.g., [12]) and working memory (e.g., [1, 13]), Sweller [9] found that problem-solving strategies may interfere with successful learning and schema acquisition. Sweller et al. [14], for example, presented participants with physics and geometry problems and varied the goal specificity. They found that nonspecific goal instructions led to faster problem-solving expertise than specific goal instructions. The reason for this effect lies in the problem-solving strategies implied by the different instructions; the goal-specific instruction engaged the study participants in the usage of means-end analysis, whereas nonspecific goal instruction eliminated the possibility of using this strategy. From a computational and structural point of view means-end analysis is a very effective problem-solving strategy resulting in far fewer dead-ends than any other strategy by breaking down a problem into a hierarchical structure of subgoals. The drawback of this strategy is that it needs working memory capacity to keep track of the hierarchical goal structure of such a magnitude that learners are not only more likely to commit mathematical errors, as shown by Sweller et al. [14] and Owen and Sweller [15], but also have little capacity for schema acquisition and learning [9].

The rationale behind these results stems from widely accepted findings about working memory capacity and its

limitations [8, 13, 16]. Current models of working memory postulate mechanisms and processes actively controlling and maintaining task relevant information. Baddeley and Hitch [1] proposed a new widely accepted functional model of working memory consisting of a central executive that controls two slave systems, the visuospatial sketchpad for visuospatial information and a phonological loop for verbal information. Both slave systems are limited in capacity and are independent of one another. If learners have to deal with concurrent tasks such as solving a problem and learning the underlying structure, their central executive is in charge of prioritizing tasks and sharing the available capacity and resources among the concurrent tasks. In the case of means-end analysis most available memory capacity is used to keep track of the problem solving process. Schema acquisition may be considered as a concurrent task to deal with in working memory. Reduced priority of this task combined with a high cognitive load task such as means-end analysis finally leads to interferences in schema acquisition and poorer reproduction performance [14].

With these impacts of cognitive load on learning in mind, Pass et al. [17] take a closer look at cognitive load and postulate three different sources of cognitive load imposed by a task on the working memory: (1) intrinsic cognitive load, (2) extraneous cognitive load, and (3) germane cognitive load. Intrinsic load is induced by the task-inherent complexity and cannot be altered by an instructional design. Task complexity is a function of element interactivity, that is, interacting elements which have to be considered at the same time in order to understand their relationship [18]. Consider, for example, the calculation of a right-angled triangle. When only one side of the triangle is taken into account, the problem will not be solved. The length of all three sides has to be considered at the same time to come up with the solution. The difficulty of the material is determined by the total number of elements that must be considered, and by the extent of their interaction [18]. Extraneous cognitive load is imposed by inappropriate design and organization of the learning material. Learners are then engaged in cognitive activity irrelevant to the solution of the problem, that is, restructuring the problem or interpreting incomprehensible instructions. Germane cognitive load is the learning relevant load that can be used for schema acquisition and metacognitive processes involved in learning. It occurs when a task is presented in a favorable way that makes it easy for learners to understand their learning processes [17]. Optimizing instructional material for cognitive load therefore involves reducing extraneous load to a minimum and maximizing germane load by encouraging the learners to use their cognitive capacities for metacognitive and learning-relevant activities.

Whereas in instructional design and learning cognitive load is often focused, CLT was seldom considered in usability or eCommerce research specifically. Reducing extraneous load may lead not only to faster learning of the site structure but also to improving users' free capacity for searching, decision making, product comparison processes, and user satisfaction in general. Conklin [19] and Evekand and Dunwoody [20], for example, described strong

interrelations between cognitive load and the lost-in-space feeling, commonly referred to as cognitive disorientation. On the one hand, cognitive load is imposed on the user's working memory by keeping track of his position in a Web site (finding out about his actual position by checking which links and menus have already been looked at and which links have still to be visited). On the other hand, the completion of complex tasks also consumes cognitive resources and therefore may reduce capacity to keep track of the navigational position, thus producing disorientation. Feelings of disorientation may induce additional cognitive load and diminish user satisfaction and motivation to visit the site again [20].

Related to this topic, Katz and Byrne [21] suggested the use of the local search function on a Web site which would depend on multiple factors such as personal preference, the Web site, and menu structure. In observing participants while they had to locate a list of items in different Web shops, Métrailler et al. [22] found that the success rate was higher when participants used the site search function than when they were browsing the menu; furthermore, browsing processes took longer than searching processes. It could be argued that users who looked for the products with the search function were able to locate a product directly and they did not need to spend their resources on learning the structure of the menu and keeping track of their position.

Applying CLT [9] to the domain of usability research, Chevalier and Kicka [23] investigated cognitive load issues during information retrieval on an ergonomic and a non-ergonomic Web site. Contrary to the principles of CLT and literature of problem solving (e.g., [10, 11]), which suggest, that expert users should experience lower cognitive load as a result of chunking, automation processes, and forward problem solving, they could not find any differences in cognitive load between professional designers, experienced users, and novices. To account for these findings, the authors pointed out that experts and users did not handle the page in the same way but they were not able to provide a more detailed analysis of these outcomes. Surprisingly, they found the ergonomic site consumed more cognitive resources than the nonergonomic site for the users, but not for the expert group. This finding is explained by users being more able to focus on the relevant task when interacting with the ergonomic site than when working with the nonergonomic site.

**2.1.2. Cognitive Load Measurement.** The methods that are used to measure cognitive load can be classified based on the two dimensions of *objectivity* (subjective or objective) and *causal relationship* (direct or indirect) [24]. Objectivity differentiates between self-reported data or subjective impressions on the one hand and objective observations of behavior, performance, or physiological reactions on the other hand. Causal relationship reflects the type of relation between cognitive load and the phenomenon observed by the measure. A direct link, for example, exists between cognitive load and the difficulty of the learning materials because difficulty is a direct result of intrinsic and extraneous load of the material. An indirect relationship exists between cognitive load and the frequency of navigation errors. Navigation

TABLE 1: Classification of methods for measuring cognitive load based on objectivity and causal relationship (adapted from [24]).

		Causal relationship	
Subjective	Self-reported	Self-reported	Self-reported
	invested mental effort		stress level
Objective	Physiological	Brain activity	measures
	behavioral measures; learning outcome measures	(e.g., fMRI); dual-task performance	

errors may be caused by an incomplete mental model of the Web site, which itself may be due to high cognitive load [24]. This leads to four different categories of cognitive load measurement methods: (1) indirect and subjective, (2) direct and subjective, (3) indirect and objective, and (4) direct and objective (cf. Table 1).

Indirect and subjective methods that are frequently used in instructional research assess learners' invested mental effort with posttreatment questionnaires (see, [25]). NASA-TLX [26] is an example of an indirect, subjective assessment method of mental workload which has its origins in research on mental workload in aviation and cockpit design. NASA-TLX is a questionnaire consisting of 6 items including mental, physical, and temporal demands as well as performance, effort, and frustration level. NASA-TLX provides a very simple and quick technique for operator workload estimation with generic items that can be applied to any domain [27]. Still, certain disadvantages are inherent in this category of measurement methods, including a subjective perception of effort which is only assessed after task completion. Furthermore, the connection between subjective and effective workload is unclear.

Direct, subjective measures involve rating of the difficulty of the material which directly related to the cognitive load imposed.

Indirect, objective measures include measuring performance outcomes such as task completion time (TCT) or learning outcomes. Here, the instructional design is usually varied whereas for the different conditions the same materials are used. Thus the intrinsic load of the material is held constant whereas extraneous load (and germane load) is investigated (e.g., [28]). Performance on these tasks may then give a hint to the cognitive load imposed.

Direct, objective measures involve dual-task methodologies and functional brain imaging. In functional brain imaging (e.g., [29–31]) brain activation is measured while working memory tasks are being carried out and may be considered a direct indication of cognitive load. The downside is that brain imaging still has low practicability for designers and engineers. Dual-task methodology is a second direct and objective way of assessing cognitive load that has been widely used in working memory and attention research (e.g., [13, 29, 32]). It is based on the flexible allocation of cognitive resources to different tasks using the same memory structures. It is assumed that concurrent verbal tasks share

the limited capacity of the phonological loop, for example, and that the central executive then allocates capacity to the tasks, depending on the focus of attention. Thus, increasing allocated resources for one task will decrease the resources for the other. There are two approaches to this, both of which use dual-task methodology. The first is cognitive load manipulation, that is, imposing cognitive load with a secondary task and analyzing the effects on performance in the primary task. This methodology is often used in social cognition and marketing research (e.g., [33, 34]). The second approach of using dual-task methodology is measuring performance on the secondary task. Different versions of the primary task will then induce different amounts of cognitive load and measurably affect performance in the secondary task [24]. Brünken et al. [35] used a dual-task paradigm which can be adapted for cognitive load measurement in ecommerce applications and has rarely been used in human-computer interaction research. As a primary task, they used two different versions of a learning system consisting of 22 pages presenting information either audio-visually or visually only. The secondary task consisted of a continuous visual observation task. During execution of the primary task, study participants had to observe a small black window with a letter changing its color after a random period of 5 to 10 seconds. Participants were asked to press the space bar as fast as possible when the color changed to red. Dependent variables were reaction times for the secondary task and learning outcomes. Due to expected large individual differences, the experiment was conducted with an all within repeated measures design. As predicted with reference to CLT [9] and modality effects (i.e., influence of presentation modality on working memory performance, [36]), reaction times in the secondary task for visual-only material were significantly higher than in the audiovisual setting.

*2.1.3. Aim of this Study.* So far, the concept of cognitive load (e.g., [9]) has been presented, and different methods to measure cognitive load were introduced. Most research in cognitive load has been conducted in the domain of instruction and learning (e.g., [24]), and only a handful of studies (e.g., [37–40]) exist that address the topic of cognitive load in the context of HCI (see [41], for an extensive review). In this study, several methods of assessing cognitive load in the context of usability are used. In addition to self-reported cognitive load assessment with NASA-TLX [26], a dual-task paradigm similar to the one used by Brünken et al. [35] is adapted.

The aim of this study on the one hand is to find out whether differences in cognitive load, which result from different Web sites, can be measured with the dual-task methodology. For that purpose, data from self-reported cognitive load assessments are compared with empirical data resulting from the dual-task paradigm. On the other hand, correlations between cognitive load, search/browse preference, and user satisfaction are investigated. We assume that users who perceive cognitive load as high are rather dissatisfied with the respective Web shop.

Closely related to cognitive load is the human tendency to use heuristics and strategies in order to save memory,

cognitive resources, and time (e.g., [42, 43]). In the Internet, the users are constantly occupied with the task of spotting the relevant information, menu sections, and matching a target object to the respective category name. Katz and Byrne [21] argue that the decision whether to use site search or browsing depends on several site and user specific characteristics. Besides individual differences in users' general attitudes toward using search, characteristics of the site have an effect on the decision to use search; information scent and menu structure are key factors influencing a users analysis of cost-benefit as described by [44]. We therefore consider the search function to be one of the user's opportunities of actively reducing cognitive load when browsing consumes too many resources. Thus, we assume that the users' tendency to gather information with the search function should be closely related to their experience or expectation of cognitive load.

### 3. Method

*3.1. Participants and Design.* Participants were 32 female and 3 male psychology students, mother tongue German. Their age ranged from 18 to 28 years ( $M = 21.7$ ,  $SD = 2.6$ ) with self-reported experience using the Internet between 2 and 12 years ( $M = 6.2$ ,  $SD = 2.2$ ). They participated in exchange for course credits. In this study a within-subjects design with repeated measures was used. The experiment was set up as dual-task experiment, similar to the study by Brünken et al. [35]. Participants' primary task was to search for several products in four different online bookstores. Therefore, "book store" served as an independent variable with four levels. The dependent variable for this primary task was task completion time. The secondary task consisted of a continuous visual monitoring task. Thus, dependent variables for the secondary task were reaction times (RTs) and accuracy. Furthermore, subjective mental workload and several user satisfaction measures were assessed.

#### 3.2. Materials

*3.2.1. Primary Task.* The primary task consisted of finding five predefined books on four different online bookstores (amazon.ch, buch.ch, book.ch, and buchhaus.ch). Participants were instructed not to use the search engine to ensure that all users carried out the same task and to hold intrinsic load constant. The five books were located in different categories and could be found with a minimum of two to four clicks, depending on the Web site.

*3.2.2. Secondary Task.* For the secondary task a green "R" was presented in a small window at the right side of the browser window, as shown in Figure 1. The character changed color to red randomly between 7 and 17 seconds after presentation. Participants had to press the left control button on the keyboard as fast as possible when the color changed. The time lapse between color change and reaction (i.e., RT) was measured and saved into a logfile. When three seconds passed without response after the change of the color the initial state was restored and color switched back to green.

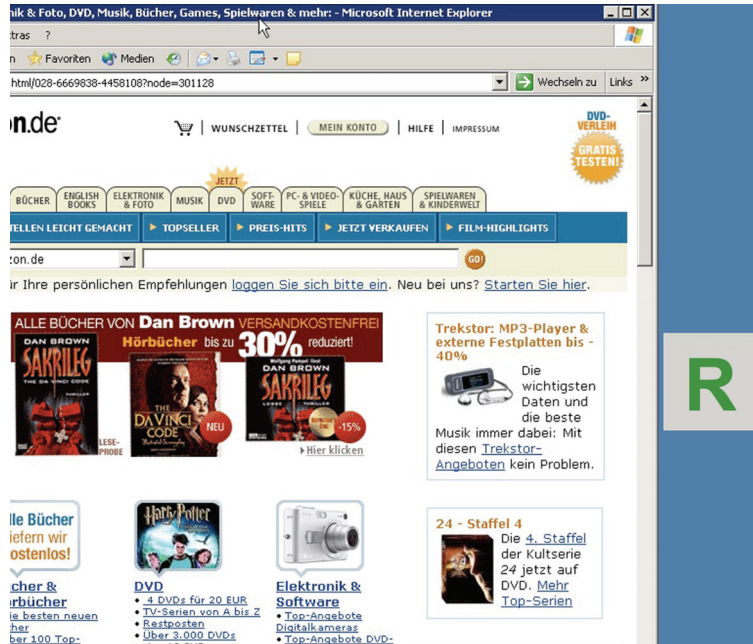


FIGURE 1: Screenshot of the materials used. On the left hand side, the browser window displaying on of the four bookstores is shown. On the right hand side, the window presenting the secondary task can be seen.

3.2.3. *Subjective Mental Workload.* Subjective mental workload was measured using an adapted and translated German version of the NASA Task Load Index (NASA-TLX, [26]), using 10-point scales. It asked the participants the following questions: (1) how much mental and perceptual activity was required? (Mental Demand); (2) how much time pressure did you feel due to the pace at which the tasks occurred? (Temporal Demand); (3) how hard did you have to work to accomplish your level of performance? (Effort) (4); how successful do you think you were in accomplishing the task? (Performance); (5) how discouraged, irritated, or annoyed did you feel during the task? (Frustration Level). In its standard version NASA-TLX consists of an additional item assessing the physical demands of the task, an item that stems from the cockpit and aviation research origins of this questionnaire. This item was discarded because of minor practical relevance to this study; due to the fact that using the search facility was disallowed and reaction to the secondary task was done by pressing the control button with the left hand, physically challenging keyboard mouse switches were not involved. The mouse actions needed to navigate through the page were not assumed to produce noticeable differences in this item between the shops. Small differences would have been overshadowed by motor conflicts between secondary task fulfillment and mouse navigation.

3.2.4. *Search/Browse Preference and User Satisfaction.* As mentioned earlier, the participants were instructed not to use the search engine. After task completion participants were asked to indicate whether they would have preferred to use the Site search function instead of navigating in the particular shop on a 10-point scale with possible answers between 1

(not at all) and 10 (very much). In addition, general user satisfaction has been measured using a 10-point scale ranging from 0 (very dissatisfied) to 10 (very satisfied).

3.3. *Procedure.* Test sessions took place in a usability laboratory of the Department of Psychology. The laboratory was equipped with a 2.3 GHz Pentium IV computer with a 19" display running a resolution of  $1152 \times 864$  pixels. Each participant was tested individually. The experimenter was in the observation room during the test sessions and followed participants' behavior via observation cam and speakers. The sessions began with a short instruction given by the experimenter, and participants were asked to give their informed consent. Then participants had to start searching for the products in one of the four online bookstores, and at the same time they had to monitor the green "R" at the right side of the browser window. Afterwards they completed the NASA-TLX questionnaire and answered to the search/browse preference and satisfaction items. This procedure was repeated for the remaining three shops. The shop sequence was randomized in order to start each session with a different shop, countering exercise effects. After finishing the trials, participants filled out a short demographic questionnaire. The entire procedure took approximately 45 minutes.

## 4. Results

Primary task performance was measured as task completion time (TCT) for each of the four different shops. 2 Participants quit the test session after 3 shops for private reasons. Data from these participants were included in the

analysis. Participants managed to find all items. Individual average RTs and accuracy measures were calculated for each participant for each shop condition. All RTs above 1200 milliseconds were scored as a miss. The individual time participants spent on each shop condition resulted in different numbers of required dual-task reactions, that is, when it took a participant a long time to find a product, more secondary task reactions were required. Therefore, accuracy was computed as correct reactions divided by the total number of reactions required by the secondary task in each condition.

*4.1. Primary Task Performance.* The descriptive values for task completion time (TCT) are presented in Table 2. An ANOVA for repeated measures (RM-ANOVA) shows that there are significant differences between the four conditions for TCT,  $F(1, 3) = 19, 8, P = .001$ . Participants on average spent most time on amazon.ch to find the required 5 items, whereas they were twice as fast in finding the books on books.ch.

*4.2. Secondary Task Performance.* The descriptive values for the secondary task performance measures (RTs and accuracy) are also presented in Table 2. Analyzing data with Repeated Measures ANOVA we found, contrary to our initial hypothesis, that RTs on the secondary task did not differ between the four shops  $F(1, 3) = .93, P = .43$ . Marginally significant differences could be found for accuracy,  $F(1, 3) = 2, 45, P = .07$ . To further analyze accuracy data, single comparisons between the four shops were conducted. Paired samples *t*-tests revealed significant differences in accuracy, namely between buch.ch and buchhaus.ch,  $t = 2.65, P < .01$ . Marginally significant differences resulted from comparing amazon.ch with buchhaus.ch,  $t = 1.90, P = .065$  and buch.ch with books.ch,  $t = 2.02, P = .053$ . Participants accuracy was generally low, solving the secondary task in about half of the trails correctly.

*4.3. Subjective Mental Workload.* Means and standard deviations for the NASA-TLX measure are also shown in Table 2. Using an ANOVA for repeated measures to compare NASA-TLX data from the four shops, significant differences in the subjective assessment of cognitive load were revealed,  $F(1, 3) = 14.0, P < .001$ . Again, amazon.ch showed the highest and books.ch the lowest scores meaning that the mental effort spent on Amazon was considered highest. Paired samples *t*-tests showed marginal significant differences between amazon.ch and buch.ch,  $t = 1.74, P = .09$ , and amazon.ch and buchhaus.ch,  $t = 1.82, P = .08$ , respectively. NASATLX scores for books.ch were significantly lower than for amazon.ch,  $t = 5.58, 10, P < .01$ , for buch.ch,  $t = 4.56, P < .01$ , and for buchhaus.ch,  $t = 4.34, P < .01$ , respectively. buch.ch and buchhaus.ch did not differ regarding NASA-TLX score.

*4.4. Search/Browse Preference.* Analyzing the search/browse preferences for each shop (see Table 2) using RMANOVA, significant overall differences were found  $F(1, 3) = 7.69,$

$P < .001$ . Search preferences were highest for amazon.ch and lowest for books.ch, meaning that the participants on amazon.ch would have preferred to use the site search more than on the other shops. Paired samples *t*-tests only showed significant differences between books.ch and amazon.ch  $t = 3.59, P < .01$ , and between books.ch and buchaus.ch,  $t = 2.42, P < .05$ .

*4.5. User Satisfaction.* Means and standard deviations for general user satisfaction (i.e., user satisfaction measured with the one item covering overall user satisfaction) can also be seen in Table 2. An ANOVA for repeated measures showed significant overall differences,  $F(1, 3) = 10.9, P < .001$ . Paired samples *t*-tests showed significant differences between books.ch and amazon.ch,  $t = 4.19, P < .01$ , between books.ch and buch.ch,  $t = 2.90, P < .01$ , between books.ch and buchhaus.ch,  $t = 5.72, P < .01$ , and between buch.ch and buchhaus.ch,  $t = 2.22, P < .05$ . In this measure, books.ch scored highest, followed by buch.ch, amazon.ch and buchhaus.ch, in this order.

*4.6. Correlations.* Using values that were pooled over the four shops, Pearson correlations between primary task measures (i.e., TCT), secondary task measures (i.e., RTs and accuracy), subjective mental workload (i.e., NASA-TLX score), search preferences, and general user satisfaction were calculated. Different values of *N* result from missing values in the subjective ratings as well as from the 2 participants completing only 3 of the shop trials (see Table 3). General user satisfaction significantly correlated with mean TCT, NASA-TLX score, and search preferences. Although NASA-TLX further correlated with mean TCT, secondary task accuracy, and search preferences, none of the other measures correlated with secondary task accuracy.

## 5. Discussion

We assumed that there are differences in the cognitive load that is imposed on users by searching products in four different online book shops. Cognitive load was measured with subjective (i.e., NASA-TLX; [26]) and objective (i.e., dual-task paradigm, e.g., [24]) assessment methods. Furthermore, we aimed at investigating whether these cognitive load measures are related to search preferences and user satisfaction. We assumed that users who perceive cognitive load as high are rather dissatisfied with the respective Web shop and would prefer using the site search function over the site navigation.

Participants' NASA-TLX scores differed significantly among the shops. Holding the intrinsic load of the tasks constant (i.e., every participant had to look for the same books), the resulting differences in NASA-TLX scores reflect different amounts of extraneous (information architecture, visual complexity) and germane load as defined by Sweller and Chandler [18]. In this context extraneous load is cognitive load imposed by presentation and design of the individual Web site structure. More complex and difficult structures impose more cognitive load on users' working memory. We

TABLE 2: Means and (Standard Deviations) for the six major outcome measures for each of the four shops.

	amazon.ch N = 35 M (SD)	buch.ch N = 34 M (SD)	books.ch N = 34 M (SD)	buchhaus.ch N = 35 M (SD)
Primary task				
TCT (secs)	607 (165)	472 (194)	273 (123)	422 (220)
Secondary task				
Mean RTs (msecs)	706 (101)	721 (64.5)	692 (98.3)	696 (80.7)
Accuracy (%)	46.4 (18.7)	43.8 (19.7)	51.3 (19.4)	51.1 (20.4)
Subjective measures				
NASA-TLX	28.2 (6.3)	26.2 (7.1)	19.7 (8.4)	25.5 (7.4)
Search/Browse	84.3 (2.2)	80.3 (2.0)	63.2 (2.8)	77.3 (2.7)
General satisfaction	5.41 (2.0)	5.97 (1.49)	7.28 (1.6)	5.0 (1.80)

TABLE 3: Correlations between cognitive load measures and user satisfaction.

Measure	1	2	3	4	5	6
N	138	137	137	136	137	137
1. TCT	—	.18*	-.17**	.50**	.46**	-.30**
2. Mean RTs		—	-.34**	.09	-.00	-.03
3. Accuracy			—	-.25**	-.04	.13
4. NASA-TLX score				—	.63**	-.47**
5. Search/Browse					—	-.32**
6. General satisfaction						—

\* $P < .05$ . \*\* $P < .01$

therefore consider NASA-TLX to be a valuable measure of cognitive load and mental effort. Its value is supported by strong correlations with most outcome measures used in this study. High cognitive load indicated by high NASA-TLX scores was related to longer TCT, more failures in the secondary monitor task, higher search preference, and less general satisfaction with the respective shop.

Dual-Task methodology using RTs in the secondary monitoring task revealed no differences between the four shops. Participants' poor accuracy scores in this secondary task, in which they managed to react on only about half of the color changes, may raise questions about the validity of the dual-task methodology used in this experiment. First, we suppose that the secondary task was not relevant enough for the participants to spend more effort in fulfilling both tasks at the same time. Second, users might have actively suppressed the blinking secondary task object on the right visual field. Pagendarm and Schaumburg [45], for example, found that people tend to suppress objects on the right side of a browser especially when these objects do not look like task relevant content. Thirdly, unlike in the study by Brünken et al. [35], participants were instructed to interact with the system. Participants used the mouse with their right hand to navigate through the online shop while reacting to the monitoring task by using their left hand to press a button on the keyboard. The resulting motor conflicts may also have been contributing to these results.

TCT is generally used as a measure of efficiency [46], and it differed significantly among the shops and is not

only correlated with TLX, but also with accuracy, search preference, and general satisfaction. Accuracy as a second objective cognitive load measure, derived from the secondary task, did not reveal differences. Nevertheless, it correlated with NASA-TLX and RTs in the secondary task. Although these results are not astonishing, we believe that accuracy might serve as a more sensitive measure than reaction times when interactions with the system are needed. The results showed significant search/browse preference differences between the four shops and substantial correlations to NASA-TLX score. Moreover, strong correlations with TCT and general user satisfaction indicate search/browse preference to be a promising measure for a "quick and dirty" assessment of cognitive load and user satisfaction. Katz and Byrne [21] found that the decision to use site search or navigation is influenced by menu structure, interface element prominence, information scent, and finally user dispositions. All of the former factors might contribute to extraneous load and thus influence a users preference. Undoubtedly, besides cognitive load numerous factors might have contributed to the differences found in this study. In order to make more detailed statements further research is certainly needed.

General satisfaction shows strong negative correlations with TCT as well as with NASA-TLX scores and search/browse preference. This confirms our expectations. Although the relation between satisfaction and efficiency (TCT) seems to be plausible and was expected, a recent metaanalysis by [46] showed that correlations between these two aspects of usability are generally weak. They suggest that

effectiveness, efficiency, and general satisfaction should be considered as different aspects of usability. Further research is needed to fully understand the relations between these three aspects of usability. The strong negative correlation between general satisfaction and NASA-TLX scores (meaning higher experienced cognitive load is related to weaker user satisfaction) supports the aim of reducing cognitive load in terms of enhancing user satisfaction and user experience.

Comparing the four shops, we found that books.ch scored best on most of the measures discussed above, whereas amazon.ch and buchhaus.ch shared poor results. At this point, it is not easy to specify reasons or factors that contributed to the participants' experienced cognitive load in each of the shops; the present study design does not allow for such interpretations. The shops used in this study differed very much regarding visual complexity, text usability, or scent and breadth of the information architecture. Each of these factors alone and in combination with others might increase cognitive load.

Further research with controlled experiments varying these factors and measuring cognitive load using NASA-TLX might give a clearer picture of cognitive load factors in eCommerce and Web usability in general.

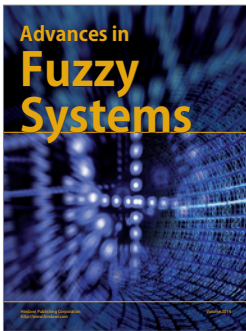
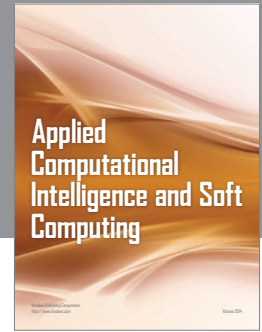
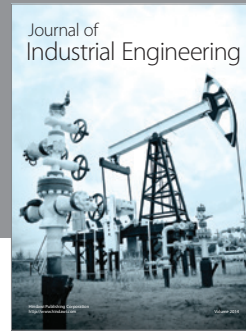
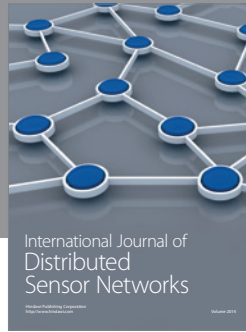
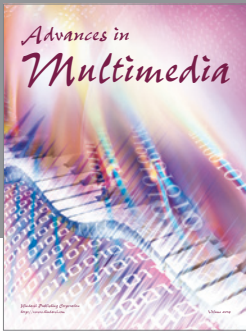
**5.1. Conclusions.** For the assessment of the usability of a computer system, NASA-TLX scores can be considered a good additional indicator of efficiency. Standard efficiency measures such as TCT as an objective measure, for example, do not take into account cognitive efficiency such as cognitive load. Objective measurement methods of cognitive load such as the dual-task methodology used in this study should be further adapted for the use with the Web and tasks needing interaction with the system. Still, it seems that NASA-TLX, TCT, and the monitoring task assess different concepts. To get a better understanding of the meaning of cognitive load in the usability context, further research could address other operationalizations of the cognitive load concept and also investigate whether they assess the same construct. A further aspect that was introduced in this study concerns search preference, which seems to be an interesting "quick and dirty" measure of complexity and cognitive load. Further research is needed if search preference is really to be used as behavioral indicator for complexity and cognitive load.

## References

- [1] A. Baddeley and G. Hitch, "Working memory," in *Recent Advances in Learning and Motivation*, G. Bower, Ed., pp. 47–90, Academic Press, New York, NY, USA, 1974.
- [2] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, pp. 152–158, Boston, Mass, USA, April 1994.
- [3] B. Shneiderman and C. Plaisant, *Designing the User Interface*, Addison-Wesley, Reading, Mass, USA, 2004.
- [4] S. Leuthold, J. A. Bargas-Avila, and K. Opwis, "Beyond web content accessibility guidelines: design of enhanced text user interfaces for blind internet users," *International Journal of Human-Computer Studies*, vol. 66, no. 4, pp. 257–270, 2008.
- [5] T. Stewart and D. Travis, "Guidelines, standards, and style guides," in *The Human-Computer Interaction Handbook*, J. Jacko and A. Sears, Eds., pp. 991–1005, Lawrence Erlbaum Associates, London, UK, 2003.
- [6] B. Shneiderman, *Designing the User Interface*, Addison-Wesley, Boston, Mass, USA, 1998.
- [7] T. Mandel, *The Elements of User Interface Design*, John Wiley & Sons, New York, NY, USA, 1997.
- [8] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
- [9] J. Sweller, "Cognitive load during problem solving: effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988.
- [10] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, "Expert and novice performance in solving physics problems," *Science*, vol. 208, no. 4450, pp. 1335–1342, 1980.
- [11] W. Chase and H. Simon, "The mind's eye in chess," in *Visual Information Processing*, pp. 215–281, Academic Press, New York, NY, USA, 1973.
- [12] J. Bruner, "The art of discovery," *Harvard Educational Review*, vol. 31, no. 1, pp. 21–32, 1961.
- [13] A. Baddeley, *Working Memory*, Oxford University Press, Oxford, UK, 1986.
- [14] J. Sweller, R. F. Mawer, and M. R. Ward, "Development of expertise in mathematical problem solving," *Journal of Experimental Psychology: General*, vol. 112, no. 4, pp. 639–661, 1983.
- [15] E. Owen and J. Sweller, "What do students learn while solving mathematics problems?" *Journal of Educational Psychology*, vol. 77, no. 3, pp. 272–284, 1985.
- [16] N. Cowan, *Working Memory Capacity*, Essays in Cognitive Psychology, Psychology Press, New York, NY, USA, 2005.
- [17] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: recent developments," *Educational Psychologist*, vol. 38, no. 1, pp. 1–4, 2003.
- [18] J. Sweller and P. Chandler, "Why some material is difficult to learn," *Cognition and Instruction*, vol. 12, no. 3, pp. 185–233, 1994.
- [19] J. Conklin, "Hypertext: an introduction and survey," *Computer*, vol. 20, no. 9, pp. 17–41, 1987.
- [20] W. P. Eveland Jr. and S. Dunwoody, "User control and structural isomorphism or disorientation and cognitive load?" *Communication Research*, vol. 28, no. 1, pp. 48–78, 2001.
- [21] M. A. Katz and M. D. Byrne, "Effects of scent and breadth on use of site-specific search on e-commerce web sites," *ACM Transactions on Computer-Human Interaction*, vol. 10, no. 3, pp. 198–220, 2003.
- [22] Y. Métrailler, J. Bargas-Avila, P. Schmutz, and K. Opwis, "Search or browse: how do people behave while navigating on e-commerce websites?" in *Proceedings of the German Cognitive Science Conference (KogWis '05)*, K. Opwis and I.-K. Penner, Eds., pp. 137–140, Schwabe, Basel, Switzerland, September 2005.
- [23] A. Chevalier and M. Kicka, "Web designers and web users: influence of the ergonomic quality of the web site on the information search," *International Journal of Human Computer Studies*, vol. 64, no. 10, pp. 1031–1048, 2006.
- [24] R. Brünken, J. L. Plass, and D. Leutner, "Direct measurement of cognitive load in multimedia learning," *Educational Psychologist*, vol. 38, no. 1, pp. 53–61, 2003.
- [25] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educational Psychologist*, vol. 38, no. 1, pp. 63–71, 2003.



- [26] S. G. Hart and L. Staveland, "Development of NASA-TLX (task load index): results of empirical and theoretical research," in *Human Mental Workload*, P. Hancock and N. Meshkati, Eds., pp. 239–250, North-Holland Press, Amsterdam, The Netherlands, 1988.
- [27] N. Stanton, P. Salmon, G. Walker, C. Baber, and D. Jenkins, *Human Factors Methods: A Practical Guide for Engineering and Design*, Ashgate, Hampshire, UK, 2005.
- [28] R. Brünken, S. Steinbacher, W. Schnotz, and D. Leutner, "Mentale Modelle und Effekte der Präsentations- und Abrufkodierbarkeit beim Lernen mit Multimedia," *Zeitschrift für Pädagogische Psychologie*, vol. 15, no. 1, pp. 16–27, 2001.
- [29] E. Awh, J. Jonides, E. E. Smith, E. H. Schumacher, R. A. Koeppel, and S. Katz, "Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography," *Psychological Science*, vol. 7, no. 1, pp. 25–31, 1996.
- [30] M. A. Just, P. A. Carpenter, and A. Miyake, "Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work," *Theoretical Issues in Ergonomics Science*, vol. 4, no. 1-2, pp. 56–88, 2003.
- [31] R. R. Whelan, "Neuroimaging of cognitive load in instructional multimedia," *Educational Research Review*, vol. 2, no. 1, pp. 1–12, 2007.
- [32] C. D. Wickens, "Processing resources in attention," in *Varieties of Attention*, R. Parasuraman and D. R. Davies, Eds., pp. 63–102, Academic Press, San Diego, Calif, USA, 1984.
- [33] D. T. Gilbert, B. W. Pelham, and D. S. Krull, "On cognitive busyness: when person perceivers meet persons perceived," *Journal of Personality and Social Psychology*, vol. 54, no. 5, pp. 733–740, 1988.
- [34] B. Shiv and A. Fedorikhin, "Spontaneous versus controlled influences of stimulus-based affect on choice behavior," *Organizational Behavior and Human Decision Processes*, vol. 87, no. 2, pp. 342–370, 2002.
- [35] R. Brünken, S. Steinbacher, J. L. Plass, and D. Leutner, "Assessment of cognitive load in multimedia learning using dual-task methodology," *Experimental Psychology*, vol. 49, no. 2, pp. 109–119, 2002.
- [36] P. Chandler and J. Sweller, "Cognitive load theory and the format of instruction," *Cognition and Instruction*, vol. 8, no. 4, pp. 293–332, 1991.
- [37] S. Dennis, R. McArthur, and P. Bruza, "Searching the world wide web made easy? The cognitive load imposed by query refinement mechanisms," in *Proceedings of the 3rd Australian Document Computing Symposium (ADCS '98)*, pp. 65–71, Sydney, Australia, August 1998.
- [38] S. Joseph, J. Yukawa, D. Suthers, and V. Harada, "Searching emergent vocabularies: exploring methods to reduce cognitive load during web navigation and resource contribution," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06)*, vol. 7, p. 154a, Kauai, Hawaii, USA, January 2006.
- [39] J. Kjeldskov and J. Stage, "New techniques for usability evaluation of mobile systems," *International Journal of Human Computer Studies*, vol. 60, no. 5-6, pp. 599–620, 2004.
- [40] Y. L. Theng, M. Jones, and H. W. Thimbleby, "Reducing information overload: a comparative study of hypertext systems," in *Proceedings of the IEE Colloquium on Information Overload*, pp. 1–4, London, UK, November 1995.
- [41] K. Hornbæk, "Current practice in measuring usability: challenges to usability studies and research," *International Journal of Human Computer Studies*, vol. 64, no. 2, pp. 79–102, 2006.
- [42] S. Chaiken, A. Liberman, and A. H. Eagly, "Heuristic and systematic processing within and beyond the persuasion context," in *Unintended Thought*, J. Uleman and J. Bargh, Eds., pp. 212–252, Guilford Press, New York, NY, USA, 1989.
- [43] G. Gigerenzer, "Fast and frugal heuristics: the tools of bounded rationality," in *Blackwell Handbook of Judgment and Decision Making*, D. J. Koehler and N. Harvey, Eds., pp. 62–88, Blackwell, Oxford, UK, 2004.
- [44] P. Pirolli and S. Card, "Information foraging," *Psychological Review*, vol. 106, no. 4, pp. 643–675, 1999.
- [45] M. Pagendam and H. Schaumburg, "Why are users banner-blind? The impact of navigation style on the perception of web banners," *Journal of Digital Information*, vol. 2, no. 1, 2001.
- [46] K. Hornbæk and E. L.-C. Law, "Meta-analysis of correlations among usability measures," in *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pp. 617–626, San Jose, Calif, USA, April 2007.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

