Towards Cognitive Load Theory as Guideline for Instructional Design in Science Education

Barbara Meissner[1,*] & Franz X. Bogner[1]

[1]University of Bayreuth, Chair of Biology Education, 95440 Bayreuth, Germany

*Corresponding author: University of Bayreuth, Chair of Biology Education, 95440 Bayreuth, Germany
Tel: 49-(0)921-552-591   E-mail: barbara.meissner@uni-bayreuth.de

Abstract

We applied cognitive load theory in an heuristic out-of-school science lesson. The lesson comprises experiments concerning major attributes of NaCl and was designed for 5th to 8th grade students. Our interest focused on wether cognitive load theory provides sufficient guidelines for instructional design in the field of heuristic science education. We extracted student clusters derived from pre-knowledge and learning success. We characterised students, based on cognitive achievement, mental effort, and instructional efficiency. Cluster analyses revealed three student clusters with quite satisfying results. Two further clusters showed improvable results, two showed no learning success, which may point to difficulties in coping with the learning setting. Motivational characterisation will refine the results, and may confirm starting points to advance cognitive load theory in heuristic science education.

Keywords: *cognitive load; instructional efficiency; heuristic science education; out-of-school*

1. Introduction and Theoretical Background of Cognitive Load Theory

To describe the status of working memory in learning situations, the cognitive load theory (CLT) uses the concept of element interactivity. CLT differentiates between element interactivity caused by material to be learned (imposing intrinsic cognitive load on a learner's working memory), and element interactivity caused by processing information that is not relevant for learning (imposing extraneous cognitive load). Thus, intrinsic cognitive load depends on the difficulty and complexity of a task, while extraneous cognitive load results from instructional design. An additional component, germane cognitive load, is defined as referring "to the working memory resources that the learner devotes to dealing with the intrinsic cognitive load" (Sweller 2010, p. 126). (Sweller 2010)

On the basis of CLT, many instructional designs have been monitored with regard to learning effects. As a result, clear guidelines for instructional design have been developed (e.g. Sweller, Van Merriënboer, & Paas, 1998; Sweller, 2010), and CLT has proved to be a valuable theory of instruction (Ozcinar, 2009, Sweller & Chandler, 1991; Paas, Van Gog, & Sweller, 2010). We applied its principles for instructional design to a heuristic science lesson at an out-of-school learning setting. We were interested in the following questions: In how far is the lesson adapted to each student's individual requirements? What should further improvements focus on? For this purpose, we analyzed student characteristics to obtain a profile of our lesson.

Our project combined curricular topics and contents with out-of-school experiences in school life – an approach often required by science education research (Braund & Reiss, 2006; Hofstein & Rosenfeld, 1996), but also criticised as problematic (Kirschner, Sweller, & Clark, 2006). On the one hand, capacity of working memory is constrained (Baddeley, 1992), on the other hand, students are expected to demonstrate clear cognitive achievement after participation in a curriculum-based outreach project. We developed an interactive out-of-school lesson concerning major attributes of salt (NaCl), following the principles of CLT. We characterized students to examine the value of CLT as a guideline for the design of heuristic out-of-school settings.

The application of CLT to instructional design aims to optimise cognitive load. Design of demanding tasks requires (i) adequate levels of intrinsic cognitive load, (ii) reduction of extraneous cognitive load, and (iii) enhancement of germane cognitive load.

(i) There are different approaches to align intrinsic cognitive load of complex tasks (Ayres, 2006). Many recommendations are about restructuring a given task into smaller, less complex units. Intrinsic cognitive load is not varied directly, but several less intricate tasks are developed (Van Merriënboer, Clark, & de Croock, 2002). Direct reduction of intrinsic cognitive load can take place through a simplification of a complex task followed by the presentation of more complex versions step by step (Pollock, Chandler, & Sweller, 2002; Van Merriënboer et al. 2002; Van Merriënboer, Kester, & Paas, 2006).

(ii) For a reduction of extraneous cognitive load, many approaches are well-known, of which we employ some relevant for our study (cf. Sweller et al. 1998; Sweller 2010): Split-attention effects (Sweller, Chandler, Tierney, & Cooper, 1990) occur if learners have to keep in mind different issues simultaneously: Mental integration of information from different sources increases element interactivity in working memory. Split-attention effects are reduced if information is given in condensed rather than separated mode (e.g. comments integrated into a figure). Redundancy effects (Chandler & Sweller 1991) occur if a task comprises much information unnecessary for understanding: Learners have to invest working memory capacity to process redundant information, which results in unnecessarily interacting elements. This effect also includes the expertise-reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) as a learner's previous knowledge and expertise determine wether certain information turns out to be redundant or not: High expertise learners may be confronted with more redundant information than novices. The problem-completion effect is very similar to the worked-example effect: If a task already provides a framework of solution steps (completion problems) or the complete solution (worked examples), learners do not have to apply the very demanding means-end strategy (Kalyuga, Chandler, Tuovinen, & Sweller, 2001) for problem solving, which results in reduced element interactivity.

(iii) To enhance learning processes (i.e. to foster germane cognitive load), motivation plays an important role: It is the learner who decides wether to invest working memory capacity for learning processes (Van Merriënboer, Schuurman, De Croock, & Paas, 2002; Van Merriënboer, Jelsma, & Paas,1992). Learners need adequate stimulation in order to expend working memory capacity for learning processes (Schnotz & Kürschner, 2007). For this purpose, tasks of high variability and an appropriate level of guidance are known to be advantageous (Van Merriënboer et al., 2006). High variability enables learners to become familiar with the conditions under which certain methods can be applied. Thus, high variability leads to more complex cognitive schemata and facilitates transfer of knowledge (Paas & Van Merriënboer, 1994). Concerning adequate guidance, only the self-explanation effect has been described (Sweller, 2010): Self-explanation prompts help guide learners' thinking and help learners get aware of what they are doing. If learners need to formulate explanations they need to process information relevant for learning. Hence self-explanations enhance germane cognitive load.

2. Instruments and Methods

Considering these recommendations of the CLT, we developed an interactive out-of-school lesson suitable for a wide range of learners. After implementation, we characterised students according to cognitive parameters: We performed cluster analyses on the basis of a repeatedly applied multiple-choice knowledge test to obtain student subsamples according to the individual effectiveness of the lesson. Persistence of knowledge, mental effort ratings, and instructional efficiency scores served as cognitive parameters to characterise the different clusters. We intended to get insight in the educational value of the lesson and to find general starting points for improvement. CLT may provide an adequate repertoire of recommendations for the design of interactive outreach projects, which could enhance students' competence formation without neglecting the gain of factual knowledge.

2.1 CLT-based Lesson

Our curriculum-based out-of-school lesson is part of the educational programme of a commercial salt mine. However, in order to exclude unpredictable site effects in our analyses, the implementation took place at a neutral out-of-school learning setting with no links to the subject of salt, namely in an environmental information centre. In another study, students' cognitive outcome and emotional feedback did not differ between the learning site of a salt mine and neutral setting (Meissner & Bogner, 2011). However, the neutral surroundings guaranteed uniform test conditions.

The lesson aims to enable students to get insight in the science topic of salt (NaCl) and to gain first impressions of working with laboratory equipment. The workstations of our out-of-school lesson incorporate five experiments that illustrate important attributes of salt on a basic level. They cover the issues 'freezing point depression', 'electric conductivity', 'endothermic solvation processes', 'density increase', and 'osmotic activity'. Students worked together in small groups. Group composition was left to students' choice (Ciani, Summers, Easter, & Sheldon, 2008).

Referring to Sweller (2004), we assigned to the instructional materials the function of a central executive, as they helped order and structure information and activities: Instructional guidelines contained illustrated step-by-step instructions offering appropriate guidance and facilitating hands-on activities. Additionally, to ensure that students reached the educational objectives, each student was provided with a workbook containing tasks to document observations, display results, and draw conclusions.

The preconditions defined a learning situation with high extraneous cognitive load: The coordination of instructions, workbook tasks, and equipment implied a split-attention effect caused by "spatially separated" (Schnotz & Kürschner, 2007, p. 471) sources of information. Furthermore, the novelty of the learning place was supposed to produce additional cognitive load that would not contribute to learning. We designed instructional guidelines and workbook tasks according to CLT principles to compensate for the demanding setting.

2.1.1 Adequate Level of Intrinsic Cognitive Load

As the extraneous cognitive load of the lesson was very high we reduced intrinsic cognitive load to reach adequately challenging learning conditions (Paas, Van Merriënboer, & Adam, 2004). Students did not have any chemical lessons, yet, and had no chemical pre-knowledge as well as experiences with laboratory devices. Thus, we simplified the complex tasks, limited the experiments to phenomenological descriptions of the effects, and provided clear methodical guidelines. We used the terms 'salt-particles' and 'water-particles' in our explanations, as the concept of 'particles' is already part of 4th grade curricula.

2.1.2 Reduction of Extraneous Cognitive Load

As illustrated in Fig. 1, we placed illustrations of the instructional guidelines beside the corresponding text, and structured the text in subsections according to working steps. Consequently, we facilitated performance as students could (a) read the instruction for a certain working step and look at supporting illustrations (b) perform the required activities (c) easily retrieve in the guidelines the step to be taken next. In this fashion we reduced split-attention effects: Element interactivity was decreased as students were instructed to keep in mind and process one step by another.

We constructed workbook tasks in the form of completion problems (Van Gog & Paas, 2008). That is, students completed prestructured tables, texts, pictures, and so on (cf. Fig. 1). Therefore, a clear guideline facilitated careful handling of the tasks.

We excluded redundant information and concentrated on the concise description of working steps in the instructional guidelines, and precisely formulated tasks in the workbook. For each workstation, we displayed some interesting additional information separated from the tasks for students to read optionally.

Figure 1: Excerpts of Instructional Guidelines (left) and Workbook (right), Each referring to workstation 3
(Introductory title page not shown)

2.1.3 Fostering Germane Cognitive Load

The nature of the lesson itself implied high task variability: Each workstation presented a similar problem – performing an experiment – under an individual surface story – the subject of the workstation (cf. Van Merriënboer, Kester, & Paas, 2006). Workbook tasks prompted group discussions as students were asked to document their individual performance outcomes. Furthermore, introductory labels were designed to provoke curiosity and stimulate students' discussions about possible effects or explanations.

Fostering germane cognitive load means fostering students' motivation. Hence we included main statements of intrinsic motivation research (e.g. Reeve, 1996; Ryan & Deci, 2000) and considered students' basic needs for autonomy, relatedness, and competence. (cf. Meissner & Bogner, 2013)

2.2 Target-Group and Participants

As in the salt mine typical school visitors' age and school type are wide spread, we designed a project for 5th to 8th graders of various stratification levels. 17 classes from 10 schools, a total of 276 students (M_{age} = 11.6, SD = 1.6), participated in the study. The sample included students of two age groups and two stratification levels (cf. Table 1).

Table 1: Sample Description

	Subsample		
	1	2	3
n	91	104	81
Stratification level	Low	High	Low
Age group	10-12	10-12	13-15
Male/female	54 / 37	30 / 74	44 / 37

As students of our target group were of various ages and stratification levels, we assumed a wide range of expertise. Accordingly, for students with low expertise, the mental load of the tasks would be almost too high to be performed completely without help. On the other hand, highly experienced students might find tasks easy, and might even be able to solve more complex tasks without help. To obtain a well-balanced situation, we tried to avoid non-taxing demands upon higher expertise students without overcharging less expertised students.

2.3 Knowledge Tests as Instrument to Value Effectiveness of the Lesson

As a wide range of learners participated in the study, we were interested in specific student characteristics. We chose students' performance on a knowledge test to describe the individual effectiveness of the lesson in terms of students' "having the power to produce, or producing, a desired result" (Chambers 21st Century Dictionary, 2010). The knowledge test comprised 13 multiple-choice items concerning major outcomes of the workstations. Examples of knowledge test items are listed in Table 2. We applied the knowledge test one week before (pretest; KT1), immediately after (posttest; KT2), and six weeks after the lesson (retention test; KT3). Each time, the order of questions and the order of distractors within each question were varied to obviate test effects.

Table 2: Examples of Knowledge Test Items

Category	Example
Effect	Which of these conducts electricity the best?
	Pure salt / Pure water / Rock salt / Saltwater [correct]
Device	What are binoculars used for? To:
	see things amplified [correct] / dissolve substances / measure indoor and outdoor temperature simultaneously / gauge objects exactly

2.3.1 Knowledge Test Piloting

The knowledge test was pilot-tested with 109 5th grade students (high stratification level) who filled in the pretest one week before and the posttest immediately after the lesson (Cronbach's alpha = .72). The composition of knowledge test items proved to be adequate for students, as the amount of correct answers per item (difficulty index) ranged between 22 % and 84 % in the pretest (i.e. students had not dealt with the material before; cf. Fig. 2), and between 45 % and 91 % in the posttest (i.e. after a subject-specific lesson). Corrected item-total correlation (discrimination index) was .357 on average in the posttest (cf. Fig. 2). Thus, items were appropriate to differentiate

between high and low achievers after the lesson.

A control sample of 29 students (M_{age} = 13.48, SD_{age} = .63) filled in the pre- and the posttest without participating in the lesson. We used the non-parametric Wilcoxon test to compare the results. We found no significant differences between pre- and posttest scores, which indicated that no test effect occurred: students who simply filled in the knowledge tests without any treatment did not gain any knowledge.
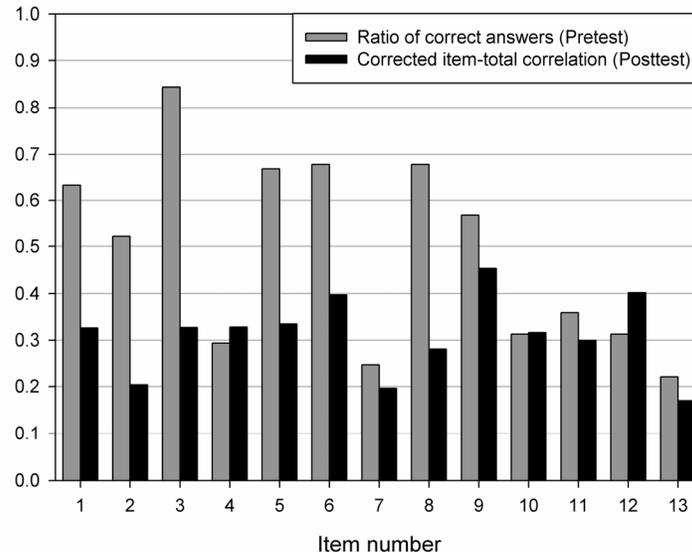

Figure 2: Ratio of Correct Answers in the Pretest, and Corrected Item-Total Correlation in the Posttest for Each Knowledge-test Item

2.3.2 Pre-Knowledge and Cognitive Achievement

Repeated application of the knowledge test enabled us to estimate students' pre-knowledge and cognitive achievement. These values indicated the effectiveness of the lesson. We took the sum-scores of KT1 (pretest; applied one week before the lesson) as indicators of students' previous knowledge. A weighted difference between sum-scores of KT2/3 and KT1 (Scharfenberg, Bogner, & Klautke, 2007) yielded short-term/long-term cognitive achievement scores. For instance, we calculated long-term cognitive achievement scores as follows:

(sum-score KT3 – sum-score KT1) * (sum-score KT3 / total number of items)

The quotient 'learner's sum-score in KT3/total number of knowledge test items' was used to diminish ceiling effects caused by the restricted number of items (cf. Scharfenberg et al., 2007).

We chose long-term cognitive achievement scores as an index of learning success, as the process of learning is defined as the result of changes in long- rather than short-term memory (e.g. Schnotz & Kürschner, 2007). As data were not normally distributed, we used the Wilcoxon test to compare knowledge test results of the whole sample, as well as short- and long-term results of each cluster.

2.4 Formation of Clusters

As variables of cluster analyses we chose the amount of pre-knowledge, well-known to be a crucial student characteristic, and learning success (long-term cognitive achievement) as the main cognitive outcome. Using these key cognitive parameters, we intended the student clusters to reflect the individual effectiveness of the lesson for different groups of students.

To guarantee equidistance of achievement score data we divided students into two groups: one group without measurable cognitive achievement (learning success = 0; n = 63), and one group with measurable cognitive achievement (learning success > 0; n = 213). On the basis of data of students with learning success > 0, we conducted two cluster analyses with pre-knowledge and learning success as variables: We compared squared Euclidian distances of different solutions yielded by hierarchical cluster analysis (Ward method) to estimate possible numbers of clusters. These tentative solutions were revised by cluster centers analysis (k-means method). To compare the cluster composition of corresponding solutions, we calculated Pearson's Contingency Coefficient c and corrected contingency coefficient c_{corr} = c/c_{max} (c_{max} = sqrt((n-1) / n) where n is the number of clusters).

Context-related comparison was applied to the results. We estimated cluster heterogeneity and used one of the methods suggested by Bergmann, Magnusson, and El Khouri (2003) to value the "percentage of the total error sum of squares 'explained' by the classification" (Bergmann et al., 2003, p. 99). We described the clusters of the final solution in relation to the quartiles (Q) of the pre-knowledge and learning success scores of the whole sample (cf. Figs. 4 and 5): Scores < Q1 were labelled *low*, scores ≥ Q1 and < Q3 *medium*, and scores ≥ Q3 *high*.

According to this allocation, we divided the group of students without measurable cognitive achievement (learning success = 0) into three subsamples that corresponded to students' different levels of pre-knowledge. To facilitate understanding, these subsamples will also be called '(artificial) clusters' in the following.

2.5 Self-rated Mental Effort

To obtain data for the calculation of instructional efficiency, and for an overall estimation of "resource requirements" (Gopher & Braune, 1984, p. 529) of the lesson, we asked students to rate their perceived mental effort (ME). We applied the one-item mental effort self-rating scale proposed by Paas (1992) and confirmed to be applicable by Sweller (2010). The scale is based on the perceived difficulty scale developed by Bratfisch, Borg, and Dornic (1972). We applied the scale against the background that self-ratings presumably allow deeper insight in student characteristics than objective (e.g. physiological) measurements (cf. Bratfisch et al. 1972): Self-perception would be more relevant for "a person's feelings, attitudes, motivation, etc." (Bratfisch et al. 1972, p. 1).

We applied the ME self-rating scale five times during the lesson: Each time students had completed a workstation, they rated their amount of invested ME on a symmetric seven-point scale. As a reference point, we set '4' as *regular science lesson*. We calculated mean scores of these five ratings to obtain the average ME invested during the lesson (Cronbach's alpha = .68). We used these average scores for the calculation of instructional efficiency scores, and as an overall estimation of "resource requirements" (Gopher & Braune, 1984, p. 529) of the lesson. For this purpose, we used the Mann-Whitney U test to compare non-normally distributed ME scores of each cluster with the corresponding residual sample.

2.6 Instructional Efficiency of the Lesson

We performed cluster analyses on the basis of students' effectiveness data (pre-knowledge, learning success) and used instructional efficiency (IE) to describe the students from a cognitive point of view. To estimate IE, we combined learning success and average mental effort during the lesson. That is, according to Van Gog and Paas (2008), we measured the instructional efficiency of the learning process itself. This combination of learning and mental effort allowed a more precise analysis than a separate analysis of both variables (Paas & Van Merriënboer, 1993; Janssen, Kirschner, Erkens, Kirschner, & Paas, 2010).

We used the method developed by Paas and Van Merriënboer (1993) to calculate IE. First, we calculated z-standardised mental effort scores (ME_z) and learning success scores (L_z) for each student. Second, IE was calculated as follows:

$IE = (L_z - ME_z) / \text{sqrt}(2)$ (Instructional efficiency score of each student)

As variables are standardised, values always are to be interpreted in relation to average. It must also be considered that instructional efficiency yields a relative score dependent on two different variables. Results therefore need careful interpretation and can not be analysed in isolation from the context. Otherwise, analysis may of course lead to wrong conclusions as indicated by de Jong (2010).

Resulting IE scores were normally distributed (Kolmogorov-Smirnov with Lilliefors correction: $p = .09$). Therefore, we used T-tests to compare IE scores for each cluster with the corresponding residual sample.

3. Results

3.1 Total Sample Results

The lesson aimed to increase students' knowledge with an appropriate amount of mental effort. In the following, we describe cognitive achievement and mental effort during the lesson with reference to the whole sample.

Fig. 3 shows the sumscores of pre-, post-, and retention tests. Results revealed significant differences between the tests (Pretest – posttest: $Z = -12.73$, $p < .001$; pretest – retention test: $Z = 5.12$, $p < .001$; posttest – retention test: $Z = 5.49$, $p < .001$).
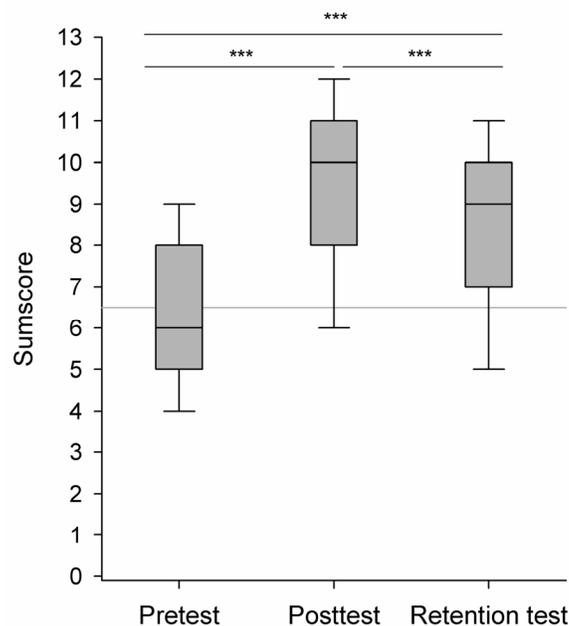
Figure 3: Results of the Knowledge Tests One Week Before (pretest), Immediately After (posttest), and Six Weeks After the Lesson (retention test) of the Whole Sample; ***: p < .001; Dotted Line Indicates Midpoint of the Scale

Mean score of the five mental effort ratings (one after each workstation) was 2.03 (*SD* = .79) on average. Scores ranged between 1 and 7, which encompasses the whole range of the scale.

3.2 Cluster Description

As a first step to characterise students, we extracted clusters based upon the effectiveness-parameters pre-knowledge and learning success. Table 3 summarises the results.

Table 3: Description of Clusters

Cluster	*n*	Heterogeneity [%]	Pre-knowledge	Learning success
I	35	18.4	Low	Low/Medium
II	28	20.2	Low	High
III	61	22.6	Medium	Low/Medium
IV	37	15.5	Medium	Medium/High
V	52	23.3	High	Low/Medium
VI	29	-[a]	High	None
VII	30	-[a]	Medium	None
VIII	4	-[a]	Low	None

'Low': scores < Q1; 'Medium': Q1 ≤ scores < Q3; 'High': scores ≥ Q3 of the whole sample

[a]Clusters VI, VII, VIII: artificial clusters without knowledge gain.

Cluster analyses (c = .84; c_{corr} = .94) resulted in five clusters (I – V) according to pre-knowledge scores (cf. Fig. 4) and learning success (cf. Fig. 5). Heterogeneity of the clusters ranged from 15 % to 23 % (cf. Table 3). There were three further, artificial clusters (VI, VII, VIII) of students without measurable learning success. As cluster VIII comprised only four students we excluded it from further analyses.
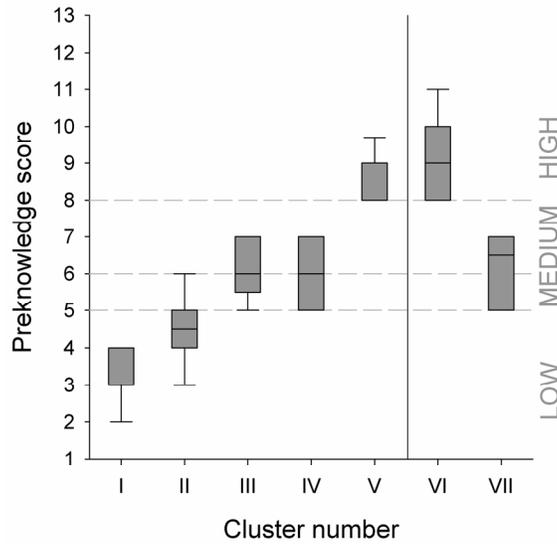
Figure 4: Pre-Knowledge Scores of the Students in the Different Clusters. Dashed Lines Indicate Quartiles of the Whole Sample

3.3 Knowledge Persistence and Mental Effort

In order to characterise the clusters, we compared short- and long-term cognitive achievement scores. Additionally, we compared mental effort scores of students of each cluster with the average score of the corresponding residual sample.

There were clear cluster differences between the results for students' short- and long-term cognitive achievement scores (Fig. 5): Students of clusters I, III, VI and VII yielded significantly lower scores for long-term than for short-term cognitive achievement (I: $Z = 3.15$, $p = .002$; III: $Z = 5.12$, $p < .001$; VI: $Z = 3.07$, $p = .002$; VII: $Z = 4.11$, $p < .001$). Short-term cognitive achievement scores additionally were spread more widely (cf. Fig. 5). Results for students of cluster II even revealed a significant increase from short- to long-term cognitive achievement scores (II: $Z = 2.86$, $p = .004$).
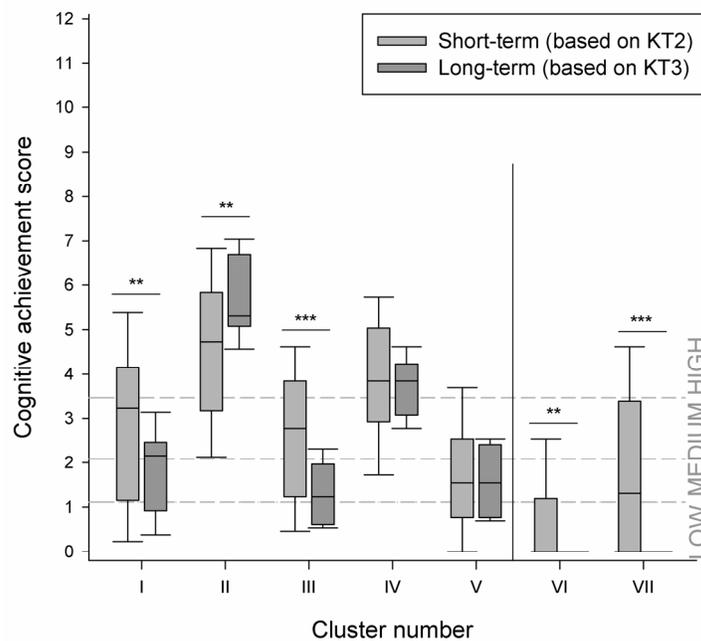

Figure 5: Short- and Long-Term Cognitive Achievement Scores. Dashed lines Indicate Quartiles of the Whole Sample; **: $p < .01$, ***: $p < .001$

Table 4 lists mean ME score for each cluster. ME scores of none of the clusters differed significantly from the mean score of the corresponding residual sample. However, clusters I, II, and VI showed slightly below-average ME. ME of clusters IV, V, and VII tended rather to be above-average, and ME of cluster III nearly equaled ME score of the total sample.

Table 4: Mean ME Scores and Standard Deviation of Each Cluster and of the Total Sample

	I	II	III	IV	V	VI	VII	Total
M	1.82	1.87	2.04	2.16	2.10	1.86	2.19	2.03
SD	.71	.70	.91	.74	.73	.60	.89	.79

3.4 Instructional Efficiency

We calculated IE scores as a combination of z-standardized ME scores and learning success scores to get insight into the characteristics of students of the different clusters. The results are shown in Table 5. IE of cluster I did not differ significantly from the residual sample, whereas clusters III, V, VI and VII revealed significantly below-average IE. Clusters II and IV showed strongly above-average IE scores.

Table 5: Instructional Efficiency Mean Score of Each Cluster and Comparison with the Corresponding Residual Sample (T-test)

	Instructional efficiency		*T*-Test		
	Cluster	Residual sample	*T*	*df*	*p*
I	0.17	0.01		*n. s.*	
II	1.63	-0.15	-10.2	270	< .001
III	-0.22	0.10	2.4	117	.018
IV	0.60	-0.06	-3.7	207	< .001
V	-0.18	0.08	2.0	105	.044
VI	-0.57	0.10	5.6	57	< .001
VII	-0.87	0.14	5.4	270	< .001

4. Discussion

We designed an interactive out-of-school lesson suitable for a wide range of learners. We characterised participating students according to cognitive parameters to estimate the educational value of the lesson and of CLT as a theory for instructional design in heuristic science education. Analysis of the results of the total sample showed that the lesson enabled students to reach clear positive cognitive achievement: Differences of both posttest and retention test compared to pre-knowledge scores were highly significant.

Mean mental effort ratings of the total sample covered the whole range of the scale from 1 to 7. However, average score of 2.03 (*SD* = .79) indicated rather low perceived mental effort. As it is not sure if mental effort self-ratings allow estimation of overall cognitive load (e.g. de Jong, 2010; Moreno, 2010), we may conclude from this result only that students did not feel cognitive overload and that we succeeded in integrating CLT principles from students' subjective point of view.

In the following, we concentrate on appraisals of each cluster (cf. Table 6). Afterwards, we outline possible conclusions that hint (a) to compliance of our lesson with CLT principles, and (b) to starting points for improvements of science instruction.

Table 6: Synthesised Cluster Description

Cluster	n	Pre-knowledge	Learning success	Short- to long-term CA[a] change[b]	Mental effort[b]	Instructional efficiency[b]
I	35	Low	Low/ medium	-**	-	~
II	28	Low	High	+**	-	+***
III	61	Medium	Low/ medium	-***	~	-*
IV	37	Medium	Medium/ high	*n.s.*	+	+***
V	52	High	Low/ medium	*n.s.*	+	-*
VI	29	High	None	-**	-	-***
VII	30	Medium	None	-***	+	-***

[a]CA = cognitive achievement. [b]'+' = increase/above-average, '-' = decrease/below-average, '~' = approximately average; **: $p < .01$, ***: $p < .001$.

4.1 Appraisal of Each Cluster

Cluster results are summarised in Table 6. Students of cluster IV yielded sustainable (i.e. no significant changes from short- to long-term results) medium to high learning success. Above-average IE confirmed the sustainability of gained knowledge and an adequate relation between mental effort and cognitive achievement. Students of cluster V with high pre-knowledge scores yielded only low/medium learning success, and below-average IE. Nevertheless, cognitive achievement was persistent, as there were no significant differences between short- and long-term cognitive achievement scores. We may therefore regard the results as satisfying, pointing to a ceiling effect: Pre-knowledge scores of students of cluster V were high. Hence, they could not have answered much more questions correctly after the lesson than before the lesson.

Students of cluster II showed high learning success, and even a significant increase from short- to long-term cognitive achievement scores. We can exclude the notion that teachers trained students before the retention test, as described by Scharfenberg, Bogner, and Klautke (2006) within science education, as each cluster comprised students from 12 to 17 different classes (cluster II: 12 classes). Any other systematic causes are implausible as these would affect either the whole sample or specific classes. Perhaps the nature of the individual learning process of these students required a kind of maturation phase until information processing and schema construction had been completed. Another reason may be that the lesson induced specific interest and, as a consequence, students' personal learning at home.

Students of cluster I with low pre-knowledge scores lost some of the shortly achieved knowledge: The low/medium learning success scores were significantly lower than short-term cognitive achievement scores. Nevertheless, IE was about average, which indicates an adequate relation between mental effort and cognitive achievement. As cognitive achievement was rather low we may conclude that students could have performed better if they had invested more mental effort. The reason they did not do so may either be motivational (students did not want to) or it may lie in insufficient guidance (students did not know how). As ME of cluster I was rather below-average, we may assume low motivation as a cause.

The two subsamples of students without measurable cognitive achievement, clusters VI and VII, and students of cluster III each revealed below-average IE and a significant decrease from short- to long-term cognitive achievement scores. The high pre-knowledge scores of students of cluster VI point to a ceiling effect that may have caused the seemingly missing cognitive achievement. Achievement scores might have been at least *low* if students had not answered at least about 2/3 of the knowledge-test questions correctly in the pretest (cf. Fig. 4). The rather below-average mental effort and IE scores of cluster VI may point to an expertise-reversal effect (e.g. Kalyuga, Ayres, Chandler, & Sweller, 2003; Schnotz, 2010; Sweller, 2010). Students of Cluster VI may have had difficulties in identifying relevant learning contents as their pre-knowledge scores were high, and they may already have developed a "knowledge-based central executive" (Sweller, 2004, p. 25) which made our central executive provided by prestructured instructional materials redundant. They may thus have estimated the lesson as not interesting, and may have opted out. However, as students of cluster V with high pre-knowledge scores showed satisfying results we can not conclude that the subject and/or the design in general were inadequate for students with high levels of previous

knowledge.

Although ME of clusters III and VII was about (cluster III) and slightly above-average (cluster VII), students' cognitive outcomes were low. Hence, students of these clusters can be supposed to have expended mental effort only to a little extent due to germane cognitive load but mainly due to extraneous cognitive load. The methods required to perform the experiments successfully may have exceeded students' cognitive abilities: To cope at the same time with workbook tasks, instructional guidelines, and equipment/devices leads to split-attention effects (Sweller, Chandler, Tierney, & Cooper, 1990; Sweller, 2010), enhanced by the novel surroundings (Orion & Hofstein, 1994). Students of clusters III and VII may have had difficulties in recognising contexts important for factual learning as they were busy with structuring their course of action. These students, 33.5 % of the total sample, may have needed more active support from the teacher. Heyne and Bogner (2011, 2009), for instance, have already applied such student-centred group work combined with individualised guidance ("immediate feedback" – Van Merriënboer, Kester, & Paas, 2006, p. 345) successfully. They demonstrated that students experiencing this kind of guidance outperformed students who worked completely on their own. Another possibility is a structured consolidation phase. We did not include such "delayed feedback" (Van Merriënboer et al., 2006, p. 345) in our lesson for organisational reasons: We assumed a consolidation phase immediately after the lesson to be pedagogically nonsensical as students already had accomplished about 90 min of physically and mentally exhausting self-guided work. On the other hand, we could not guarantee a consolidation phase in the classroom to be the same for each class as different teachers would have had to perform it.

Cluster analyses require post-hoc analyses per se. Thus, we had to draw on post-hoc statements about cognitive load although this approach has been shown to easily lead to trivial conclusions if it is not applied carefully (Schnotz & Kürschner, 2007). However, as our study does not compare the cognitive load of different instructional approaches, we may use post-hoc analysis to explain students' outcomes. In addition, our conclusions may not be overinterpreted as ME scores of the clusters did not deviate significantly from residual sample mean scores. Further motivational analyses are necessary to examine our assumptions and will be published separately.

4.2 Outline

We did not succeed in reaching all students likewise. However, we also found no hints of a systematic lack in the design of the study. We can therefore assert that the lesson was designed adequately with respect to cognitive load parameters. Instructional improvements are necessary in terms of guidance for some students, and in terms of specific motivation for others. Thus, we succeeded in identifying starting points for instructional improvement to facilitate a more effective design of out-of-school projects with adequate or at least reasonable levels of cognitive load for a wide range of low-expertise learners. Beyond a confirmation of its statements, we used CLT to identify starting points to optimise outreach projects entailing high levels of extraneous cognitive load. Further analyses of motivational parameters, for instance instructional involvement (Paas, Tuovinen, Van Merriënboer, & Darabi, 2005), are necessary to refine the results of this study, and will be published separately. Our approach may contribute to a further application of CLT in the field of science education.

A limitation of the study may lie in the fact that we did not measure unconscious learning as, for instance, methodical competencies. Analysis of the dynamics of group work with its shared cognitive capacities among group members (Kirschner, Paas, & Kirschner, 2009; Janssen, Kirschner, Erkens, Kirschner, & Paas, 2010), and training of teachers' and students' questioning strategies (Gillies & Haynes, 2010) may be helpful approaches to investigating in students' competence formation.

References

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, *20*, 281-298. http://dx.doi.org/10.1002/acp.1245

Baddeley, A. (1992). Working memory. *Science*, *255*, 556-559.

Bergmann, L., Magnusson, D., & El Khouri, B. (2003). *Studying Individual Development in an Interindividual Context: a Person-Oriented Approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Bratfisch, O., Borg, G., & Dornic, S. (1972). *Perceived item-difficulty in three tests of intellectual performance capacity*. Stockholm: Institute of Applied Psychology, University of Stockholm, Report No. 29.

Braund, M., & Reiss, M. (2006). Towards a more authentic science curriculum: the contribution of out-of-school learning. *International Journal of Science Education*, *28*(12), 1373-1388.

http://dx.doi.org/10.1080/09500690500498419

Chambers 21st Century Dictionary. http://www.chambersharrap.co.uk/chambers/features/chref/chref.py/main Chambers Publishers Ltd, (Last access April 2011).

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(3), 293-332. http://dx.doi.org/10.1207/s1532690xci0804_2

Ciani, K., Summers, J., Easter, M., & Sheldon, K. (2008). Collaborative learning and positive experiences: does letting students choose their own group matter? *Educational Psychology*, *28*(6), 627-641.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105-134. http://dx.doi.org/10.1007/s11251-009-9110-0

Gillies, R., & Haynes, M. (2011). Increasing explanatory behaviour, problem-solving, and reasoning within classes using cooperative group work. *Instructional Science*, *3*, 349-366. http://dx.doi.org/10.1007/s11251-010-9130-9

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, *26*(5), 519-532. Retrieved from http://hfs.sagepub.com/content/26/5/519

Heyne, T., & Bogner, F. (2009). Strengthening resistance self-efficacy: influence of teaching approaches and gender on different consumption groups. *Journal of Drug Education 39*(4), 439-457. Retrieved from http://baywood.metapress.com/openurl.asp?genre=article&id=doi:10.2190/DE.39.4.f

Heyne, T., & Bogner, F. (2011). Guided learning at workstations about drug-prevention with low achievers in health education. *Journal of Drug Education, 3*(1), 1-12.

Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education*, *28*, 87-112. http://dx.doi.org/10.1080/03057269608560085

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P., & Paas, F. (2010). Making the black box of collaborative learning transparent: combining process-oriented and cognitive load approaches. *Educational Psychology Review*, *22*, 139-154. http://dx.doi.org/10.1007/s10648-010-9131-x

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*(3), 579-588. http://dx.doi.org/10.1037/0022-0663.93.3.579

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31. http://dx.doi.org/10.1207/S15326985EP3801_4

Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiental, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Kirschner, F., Paas, F., & Kirschner, P. (2009). Individual and group-based learning from complex cognitive tasks: effects on retention and transfer efficiency. *Computers in Human Behavior*, *25*, 306-314. http://dx.doi.org/10.1016/j.chb.2008.12.008

Meissner, B., & Bogner, F. (2011). Enriching Students' Education Using Interactive Workstations at a Salt Mine Turned Science Center. *Journal of Chemical Education*, *88*(4), 510-515. http://dx.doi.org/10.1021/ed1006103

Meissner, B., & Bogner, F. (2013). Science teaching based on cognitive load theory: Engaged students, but cognitive deficiencies. *Studies in Educational Evaluation,* *38*(3-4), 127-134. http://dx.doi.org/10.1016/j.stueduc.2012.10.002

Moreno, R. (2010). Cognitive load theory: more food for thought. *Instructional Science*, *38*(2), 135-141. http://dx.doi.org/10.1007/s11251-009-9122-9

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching*, *31*(10), 1097-1119. http://dx.doi.org/10.1002/tea.3660311005

Ozcinar, Z. (2009). The topic of instructional design in research journals: a citation analysis for the years 1980-2008. *Australasian Journal of Educational Technology*, *25*, 559-580. Retrieved from http://www.ascilite.org.au/ajet/ajet25/ozcinar.html

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434. http://dx.doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737-743. URL: http://hfs.sagepub.com/content/35/4/737.short

Paas, F., & Van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122-133. http://dx.doi.org/10.1037/0022-0663.86.1.122

Paas, F., Van Merriënboer, J., & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419-430. http://dx.doi.org/10.2466/pms.1994.79.1.419

Paas, F., Tuovinen, J., Van Merriënboer, J., & Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: optimizing learner involvment in instruction. *Educational Technology Research and Development*, *53*(3), 25-34.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: new conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115-121. http://dx.doi.org/10.1007/s10648-010-9133-8

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61-86. http://dx.doi.org/10.1016/S0959-4752(01)00016-0

Reeve, J. (1996). *Motivating others - nurturing inner motivational resources*. Needham Heights, Massachusetts: Allyn & Bacon, (pp. 19-38 & 170-175 & 201-216).

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54-67. http://dx.doi.org/10.1006/ceps.1999.1020

Scharfenberg, F.-J., Bogner, F., & Klautke, S. (2006). The suitability of external control-groups for empirical control purposes: a cautionary story in science education research. *Electronic Journal of Science Education*, *11*(1), 22-36. Retrieved from http://ejse.southwestern.edu/article/view/7749/5516

Scharfenberg, F.-J., Bogner, F., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus - results of a teaching unit with authentic experiments. *Biochemistry and Molecular Biology Education*, *35*(1), 28-39. http://dx.doi.org/10.1002/bmb.1

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science*, *38*(3), 315-323. http://dx.doi.org/10.1007/s11251-009-9104-y

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*, 469-508. http://dx.doi.org/10.1007/s10648-007-9053-4

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*, 9-31. http://dx.doi.org/10.1023/B:TRUC.0000021808.72598.4d

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123-138. http://dx.doi.org/10.1007/s10648-010-9128-5

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, *8*(4), 351-362. http://dx.doi.org/10.1207/s1532690xci0804_5

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*, 176-192. http://dx.doi.org/10.1037/0096-3445.119.2.176

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296. http://dx.doi.org/10.1023/A:1022193728205

Van Gog, T., & Paas, F. (2008). Instructional efficiency: revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16-26. http://dx.doi.org/10.1080/00461520701756248

Van Merriënboer, J., Jelsma, O., & Paas, F. (1992). Training for reflective expertise: a four-component instructional design model for complex cognitive skills. *Educational Technology: Research and Development*, *40*(2), 23-43. http://dx.doi.org/10.1007/BF02297047

Van Merriënboer, J., Clark, R., & de Croock, M. (2002). Blueprints for complex learning: The 4C/ID-Model. *Educational Technology, Research, & Development*, *50*(2), 39-64. http://dx.doi.org/10.1007/BF02504993

Van Merriënboer, J., Schuurman, J., De Croock, M., & Paas, F. (2002). Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*, 11-37. http://dx.doi.org/10.1016/S0959-4752(01)00020-2

Van Merriënboer, J., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology*, *20*, 343-352. http://dx.doi.org/10.1002/acp.1250