Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 64 (2006) 1061–1070
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Revising (multi-) media learning principles by applying a differentiated knowledge concept

S. Guttormsen Schär[a,*], J. Kaiser[b]

[a] Institute for Medical Education, Faculty of Medicine, Inselspital 37a, CH-3010 Bern, Switzerland
[b] Business Information Systems, University of Cooperative Education Ravensburg, Germany

Received 13 April 2005; received in revised form 30 May 2006; accepted 30 May 2006
Communicated by K.S. Severinson Eklundh
Available online 28 July 2006

Abstract

This paper reports on a study investigating the effect of single-media and multimedia presentations on the resulting knowledge. First, this study investigated the stability of established multimedia learning principles by measuring acquired knowledge in different ways. Second, we aimed at testing the effect of cognitive load induced by various media combinations. Third and most important, we investigated the effect of various media combinations on the resulting kind of knowledge using a differentiated knowledge concept. Our study delivered interesting insight about multimedia effects, suggesting that the effect of (multi-) media must be evaluated with regard to the learning goals. Students do not either learn or not learn. Rather various kinds of information can be acquired depending on the representation with verbal and visual media. Experimental research in this domain should apply a more differentiated knowledge concept than often is the case today. Our results offer an interesting differentiated view of the effect of media in this context.

Keywords: Multimedia; Cognitive load; Learning; Testing; Knowledge

1. Introduction

This study was designed to test the effects of different media combinations (text, voice and picture) by measuring various kinds of knowledge. One aim of the study was to investigate whether established media effects are supported when learning outcome is operationally defined in a differentiated way. Research in this domain has shown that it is fruitless to search for pure media effects, but rather to search for the conditions under which various media affect the learning process (e.g. Mayer and Moreno, 2002). The multimedia research has recently focused on dynamic media and is searching for didactical settings where animations consistently support learning (e.g., Hegarty et al., 2002; Tversky et al., 2002; Guttormsen Schär, 2006). However, static visual presentations continue to have educational benefits, which will survive good animations (Lewalter, 2003). The most salient benefits of static presentations have shown to be that they leave control of the learning pace to the learners; they reduce cognitive load (CL) because the learners only see one major learning step at a time; they encourage germane processing because the learners are (implicitly) encouraged to explain the changes from one frame to the next for themselves (Mayer et al., 2005). Many empirically based guidelines have emerged for how to combine media in order to support learning. Mayer and his group have published a series of principles for how to design multimedia instruction with static visual media (Mayer, 2001). Three of these principles were relevant as a frame of reference for this study:

- *The multimedia principle*: students learn better from words and pictures than from words alone.
- *The modality principle*: students learn better when words in a multimedia message are presented as spoken text rather than printed text.

*Corresponding author. Tel.: +41 31 632 35 71; fax: +41 31 632 98 71.
E-mail address: sissel.guttormsen@iml.unibe.ch (S. Guttormsen Schär).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2006.05.005

- *The redundancy principle*: students learn better from animation and spoken words than from animation, spoken words and text.

Another aim of this study was to investigate whether CL is a suitable framework for explaining eventual dependencies between media effects and knowledge form. CL is well established as a factor influencing the effect of multimedia presentations. The working memory is a bottleneck in the learning process, i.e. in the process of identifying relevant information parts in presented information and integrating these elements into a mental model. The CL Theory is concerned with the limitations of the human working-memory capacity and the measures that can be taken to promote learning by imposing adequate levels of CL to the working memory. Kirschner (2002) offers a recent review of the CL Theory.

Consequently, our aims with this study are based on two open questions: First, we wanted to investigate whether Mayer's multimedia learning principles apply in general for a differentiated span of learning outcomes? Second, we wanted to investigate whether the CL Theory is generally applicable to explain eventual relations between media representations and learning outcomes?

2. Theory

2.1. Knowledge

We have earlier defined knowledge as a desired end product of learning (Guttormsen Schär, 2006). Knowledge may have many different qualities, ranging from the ability to recall previously learned material (memory) to deep knowledge.

Memory or retention knowledge is qualitatively different from deep knowledge. In the early stage of a learning process the student perceives the presented learning content—and ideally the relevant information units are identified and remembered. Retention may comprise the recall of a wide range of material, from facts to complete theories. It is possible, however, to memorize large amounts of material without gaining deep knowledge. Furthermore, retention knowledge can result from a short learning session and can be tested straight-forward in an empirical experiment.

Deep knowledge refers to an individual's mental model as a representation of the causal structure of a system. It builds on an elaborated system of logical associations between information elements. Deep knowledge can be applied and transferred to different tasks and situations (Turban and Aronson, 1988). This is also in line with Mayer's definition of deep understanding, i.e. understanding that leads to problem-solving transfer (Mayer, 2003). The proof of deep knowledge is, that the knowledge can be transferred and applied in situations different to that in which it was originally acquired.

Deep knowledge needs time to evolve, and is not likely to result from a single learning session. An experimental setting, presenting new information can, therefore, only indirectly test deep knowledge. An important step towards deep knowledge is the ability to form logical associations between central elements of a given learning content (e.g., Sweller, 1999; Kirschner, 2002). We refer to this level of knowledge as 'inference knowledge', i.e. the ability to draw a conclusion from given evidence or facts. Such inferences represent building blocks for deep knowledge and can also result from a short learning session. Hence, we aimed at designing a learning environment where inference knowledge could be tested as the learners' ability to make correct inferences from presented facts.

2.2. Aspects of external validity of controlled multimedia learning research

Various theories of situated learning claim that the degree to which knowledge is transferable depends on the setting in which the learning took place. In the situated learning approach, knowledge and skills are learned in the contexts that reflect how knowledge is obtained and applied in everyday situations. Situated learning is a general theory of knowledge acquisition. It has been applied in the context of computer enhanced learning activities for schools that focus on problem-solving skills (Lave and Wenger, 1991). Situated cognition theory sees learning as a socio-cultural phenomenon rather than the action of an individual acquiring general information from a decontextualized body of knowledge (Kirschner and Whitson, 1997). McLellan provides a review of various perspectives on the theory (McLellan, 1995).

Our research takes a cognitive approach to the investigation of learning in that we aim at understanding the cognitive structures of knowledge as a product of a learning process. When investigating the effects of different graphical representations, the cognitive approach demands a very systematic and theory-driven approach. The theoretical approach gives a necessary frame for the interpretation of the empirical data (e.g. Scaife and Rogers, 1996; Mayer, 2003; Schnotz and Bannert, 2003; Guttormsen Schär, 2006). This approach does not oppose the situated learning approach. Rather, these two approaches follow two different but not mutually excluding research questions. We considered aspects of situated learning when designing the experimental learning setting. This is important in controlled studies, because the external validity can be affected when learning takes place in an artificial learning situation.

2.3. Cognitive load

CL is related to the limited information processing capacity of the human cognitive system. The short time memory can actively process only 7 ± 2 elements simultaneously. This limitation is a bottleneck for information processing in general and for learning in particular. Cognitive overload results as soon as the information processing demand exceeds this limit. Various cognitive strategies support an extension of the processing capacity. Multiple information elements can be coded into larger singular elements, i.e. schemata, which are categorized according to the manner in which they will be used. Once a schema has been acquired, it can undergo a process of automation in the same way as we learn a rule. The long-term memory can hold large numbers of automated schemata, and as such they influence performance once material has been properly learned (Sweller, 1999; Pollock et al., 2002).

Three types of CL can be identified: external, intrinsic and germane (Kirschner, 2002). External CL refers to the learning environment and the way the information is presented. Conventional instruction tends to expose high extraneous CL on working memory. Intrinsic CL refers to characteristics of the learning task and the effort it imposes on the learner to construct adequate and rich schemata. The intrinsic CL increases with the complexity of the learning task when the singular elements within the learning material are highly interconnected (Pollock et al., 2002). Germane CL refers to the mental effort necessary for the construction of good schemata. In order to support learning that imposes a performance change, the CL must be germane. Instructional interventions cannot change the intrinsic CL but can contribute to decrease the external CL and to optimize (and even increase) the germane CL. An increase of the germane CL can be explained as focused attention and concentration. In total, the balance of the general CL must be within the limits of the working memory.

One origin of external CL is information redundancy. One view builds on the assumption that accumulated coherent information corresponds with equally more learning. Some studies suggest that information redundancy in presentations can have a positive effect on learning, as long as the information presented is consistent and relevant (review in Large, 1996). This view has also been called the information delivery view, and was recently criticized (Mayer, 2001; Mayer and Moreno, 2002).

An information processing view on learning argues against the delivery theory's view on redundancy. Recent research supports the assumption that redundancy can induce CL through split attention (Mayer and Moreno, 1998; Mayer, 2001; Kirschner, 2002; Lowe, 2003). Several media related design principles are based on the split attention assumption. The attention-split can have different dimensions, e.g., contiguity differences between corresponding words and pictures (contiguity principle), incoherence between presented words and sound (coherence principle), modality conflicts from simultaneously demanded attention to two visual sources of information (modality principle) and the simultaneous presentation of more than two redundant sources of information (redundancy principle), (Mayer, 2001; Mayer and Moreno, 2002).

2.4. Models of learning

Mayer's multimedia learning theory presents a cognitive model of learning incorporating the presentation of information with verbal and visual media (Mayer, 2001). Besides the CL assumption, the theory builds on Paivio's dual coding assumption for information processing (Paivio, 1986). Mayer's model describes the information processing from the point where presented information (words and pictures) is perceived by the sensory modality, i.e. eyes or ears, through the organization of the information in the working memory and finally the construction into cognitive models integrated with the long-term memory. The theory suggests that we have two functionally and structurally distinct but interconnected systems for information processing, one system for the processing of non-verbal objects and events (imagery system) and one system specialized for the processing of language (verbal system).

According to the dual processing theory, split attention is related to overload on one of the initial verbal or visual information channels. This would be the case, e.g., when the visual channel must simultaneously process two information sources such as visual text and picture. The dual processing assumption and the implications for split attention was recently tested in a study (Mayer and Moreno, 1998). The study examined the effect of redundant information presented to the visual channel in comparison to presentations addressing both the visual and verbal channels in parallel. The study showed that students scored better on verbal retention, visual matching and transfer (problem solving) when the presentations simultaneously addressed the visual and verbal channel than when two redundant information sources both addressed the visual processing channel. Our study takes a similar approach, extending the testing of split attention to include both the verbal and the visual channels.

Recently (Schnotz and Bannert (2003) have criticized the parallelism between text and picture processing in Mayer's model. They claim that the mapping process between the verbal and the pictorial representations cannot be conceptualized as a structural integration, as foreseen in the dual coding theory, because the sign systems used by these two systems result in fundamental different forms of representation. In their view, only descriptive representations (verbal system) contain signs for relations, but these are not present in a depictive (visual system) representation. Furthermore, they claim that a model of multimedia learning must advocate a more differentiated view on the effects of text and picture presentation. A picture can invoke many different mental representations, even in a way that may infer with mental model construction. Moreover, earlier theories present variations of the dual coding assumption (e.g. Engelkamp, 1991; Hasebrook, 1995). Hence, research will continue to contribute to the further development of an integrated theory of multimedia learning in the next years.

2.5. Operational definitions of research questions and hypotheses

Two focused research questions were formulated based on the theory: First, does split attention on the visual channel influence learning differently than split attention on the verbal channel? One implication of a positive result would be that these two sources of split attention imply different degrees of external CL. Second, do we find the same media effects as postulated in the multimedia, modality and redundancy principle when learning outcome is measured in a differentiated way? Five hypotheses were formulated:

1. *The combination of text and voice invokes a negative learning effect*: The combination of text and voice induces split attention on the verbal processing channel caused by a conflict between the individual reading speed and the reading speed of the voice.
2. *The combination of text and picture invokes a negative learning effect*: The combination of text and picture induces split attention on the visual processing channel caused by required split attention between two concurring visual information sources.
3. *The combination of picture and voice invokes a positive learning effect*: This refers to the modality principle and postulates a general learning benefit from presentations combining voice and pictures due to a balanced load on the visual and verbal channel.
4. *Representation of information with only one medium is suboptimal to a combined presentation*: This is a generalized form of the multimedia principle, which holds that the combined presentation of words and pictures is better than words alone.
5. *The combination of text, voice and picture invokes a negative learning effect*: This refers to the redundancy principle and postulates that the combination of more than two media may hinder learning due to excessive split attention.

3. Method

3.1. Design

The experiment employed a $2 \times 2 \times (7+7)$ mixed design with the between-group factors 'groups' (two groups) and 'task-set' (two task-sets). The two groups differed in two ways: The participants in the groups were sampled from two different primary school classes, and the pupils in the two groups learned a different set of seven tasks. The within-group factors were tasks, i.e. a unique set of seven principles per group, and presentation form (seven different presentation forms): text, voice, picture, text + voice, text + picture, voice + picture, text + voice + picture. The within-group factor 'task', which refers to the singular learning tasks, served as a control for variation of task complexity. Eventual effects of the presentation forms should be independent of eventual different task complexities, as an inherent effect of the various topics. The between group design with the two task-sets should control for that information about the learning content spread among the participants. The participants in the two groups alternated on participating in the experiment. All the participants, over both groups and task-set allocation, were given three different knowledge performance tests addressing visual knowledge, verbal facts knowledge and inference knowledge. Table 1 presents all the experimental variables in an overview.

3.2. Participants

Forty-two Swiss fourth level primary school pupils participated. The mean age was 10 years. The sample

Table 1
Overview of the experimental variables

Independents			
Groups	*Task-sets*	*Presentation forms*	*Control variables*
Two primary school classes	Two sets of seven principles	Text	General memory ability
		Voice	Visual memory ability
		Picture	Gender
	(Tasks = each single principle)	Text + voice	Age
		Text + picture	Scholastic performance
		Voice + picture	
		Text + voice + picture	
Dependents			
Knowledge categories			*Other measures*
Visual knowledge			Preferences
Verbal facts knowledge			Overall satisfaction
Inference knowledge			Test time per task
			Pause time

consisted of 17 girls and 25 boys. The participants took part in the experiment as a part of regular teaching.

3.3. Learning tasks

The experimental learning task was developed from various principles. Principles are truths or laws that can incorporate both facts and concepts (Merrill, 1983; Guttormsen Schär et al., 2004; Guttormsen Schär, 2006). The participants needed to selectively perceive and to mentally integrate these facts and concepts. We selected 14 principles and designed 14 different learning tasks from the content of these principles by preparing visual and verbal presentations. An exact representation of the "real" content was secondary to the aim of presenting each principle as a collection of relatively simple logically interconnected information entities. As an example the Braille principle shall be described here. It was presented as consisting of three related information parts:

(1) The Braille is another way to read and write.
(2) Every character in the Braille code is based on an arrangement of one to six raised dots. Each dot has a numbered position in the Braille matrix.
(3) Every letter of the alphabet, punctuation marks, numbers and everything else can be represented with Braille by different combinations of the dots, e.g. the letter "D" is represented with the dots 1, 4 and 5.

The other principles were presented in a similar manner. They were related to the following issues: technology (binary encoding, trick film), social issues (population growth, division of work, teamwork), economy (recycling, energy tax, energy use, production chain, buying power, waste disposal), nature (ozone layer, greenhouse effect) and miscellaneous (Braille). In the following the 14 principles are referred to as 'tasks'. The instruction for the participants to learn these tasks was that they should try to learn the different "rules" by studying the presentations on the computer screen.

3.4. Test of learning and definitions of the knowledge categories

Three learning performance questions were produced to each of the 14 tasks, reflecting visual knowledge, verbal facts knowledge and inference knowledge. The visual knowledge questions addressed visual features of the presented pictures, e.g. colour, position, form and objects. These questions were only asked when tasks were presented with a presentation form including a picture. The verbal knowledge questions addressed the retention of particular facts related to a principle, e.g. "How many positions does the Braille matrix contain?" The inference knowledge tests addressed inferences between the entities—these were not explicitly presented in the learning session, e.g. "How does the Braille enable blind people to read"? All the questions were administered as structured interviews and were recorded.

3.5. Implementation, technical setup and randomization procedures

All the tasks were prepared in three formats. A simple text version of all the principles was first formulated and adapted to the cognitive level of the participants. The text versions were recorded as voice. The pictures were simple coloured drawings representing the text and showing all the relevant information. Irrelevant visual information that could distract the attention away from the relevant aspects was kept to a minimum (coherence principle, Mayer, 2001). The pictures did not offer any cues like arrows or legends. Hence, it was necessary to produce facts and concept information as interpretations of the pictures.

The seven presentation forms were combinations of the three basic formats text, voice and picture. In the presentation-form picture + text, the picture was presented to the right of the text. In the presentations forms voice, the voice started to play simultaneously with the presentation of the picture or the text. In presentation forms with voice only, a white field was displayed on the screen during the presentation. In the presentation forms with picture or text only, the presentation was shown in the centre of the screen.

Each task was implemented with a predefined presentation time, which was defined by the time it took to read the verbal information aloud + 50%. When the voice presentation took 90 s, e.g., the total presentation time of that task was 2 min and 15 s regardless the presentation form. A counter in the lower left screen section indicated the last 20 s of the presentation time. By the end of the presentation, the computer emitted a sound followed by a screen (2 s) to delete images on the retina.

A computer programme was designed with Macromedia Director, which controlled the presentation of the tasks and the randomization procedures. A log file recorded the participants' demography, the learning times and preferences. The tasks were permutated to the different presentation conditions by the computer programme. The participants were randomly distributed to the two groups. There were 21 participants in each group. The setting and procedure were otherwise identical between the groups. Hence, each participant in each group solved all seven tasks in the task-set and viewed all seven presentation forms. Task order, presentation form and participants' order were permutated within each group. The allocation of a presentation form to a special task was performed with a controlled randomized procedure over all the 14 tasks so that each of the seven presentation forms was applied three times to each of the 14 tasks. Hence, over the 21 participants in one group, each possible combination of task and presentation form occurred three times.

The external validity of the results was optimized through our efforts of maximizing situated learning. The participants took part in the experiment during their normal school time, and were recruited to the experiment by their teacher. He introduced their participation as a part of their regular computer training. The participants were encouraged to view their participation as a new computer experience and the test or learning aspects were not emphasized. The tasks were inspired from similar learning activities performed by the teacher.

3.6. Procedure

The following experimental procedure was followed for each participant:

- Each participant was given a standardized introduction to the experiment, explaining in general terms their task and the aim of the experiment.
- Learning session:
  1. Presentation of a task for a predefined duration.
  2. Presentation of a neutral image for 2 s.
  3. Learning performance questions.
  4. Resting time—the participants decided when to proceed with the next task.
  5. When task <7 continue by 1.
- By the end of the experiment, we asked the participants to state their preference for each of the presentation conditions on a 5-point scale. The differences between the presentation forms were shown to the participants once more by a new task and each presentation form was shown for 4–5 s.
- At last, the participants were asked to indicate on a 5-point scale the degree of satisfaction in participating in the experiment.

3.7. Control for third variables

As a control for third variables, several factors were tested one week prior to the experiment in order to avoid eventual effects of fatigue: The general memory ability was recorded with the memory test of Hamburger–Wechsler intelligence test for children (HAWIK-R, 1983). The Benton test was applied as a control for visual memory skills (Benton et al., 1994). The Benton test can in particular screen children with more than average impaired visual memory capacity, which would have been a reason to exclude participants from the analysis. The teacher delivered the school performance grades for math, German, local studies and drawing. The tests for general memory capacity and visual memory were administered some days after the experimental participation.

3.8. Scoring of the knowledge

The visual knowledge and the verbal facts knowledge measures were scored on a scale between 0 and 2, where 0 = wrong, 1 = partly correct and 2 = correct. Hence, these two knowledge categories produced non-parametric data. The learning performance for inference knowledge was scored parametrically according to the following procedure: Three independent persons were asked to evaluate the performance of the participants on the inference knowledge questions. The answers were given a score between 0% and 100% correct. They were given a suggestion for a 100% correct answer as well as alternatives for partly correct answers. A correct answer contained information about all the facts that defined the principles. A partly correct answer failed to mention all the facts or stated them only vaguely. The scoring scheme assigned different weights (in %) to the range of different answer criteria. The task of the adjudicators was then to evaluate each answer and assign a score between 0% to 100% correct. The reliability estimates between the scores were computed using an analysis of variance to estimate the reliability of the measurements (Wiener, 1971). The reliability between the scores from the three people was estimated to 0.87, which is acceptable.

4. Results

4.1. Effects of group

There was no main group effect for any of the performance variables: visual knowledge, verbal fact and inference knowledge. Group effects were found for the variables exposition time and pause time between the tasks ($p = 0.027$, $F = 4.92$ and $p = 0.011$, $f = 6.63$, respectively). Because no group effect was found for the performance variables between the groups, the data set from the 42 participants was analysed as one group for the rest of the statistical analysis.

4.2. Effects of tasks and presentation form

An ANOVA based on the mean values of performance showed a main effect of disparity for tasks ($F = 8{,}56$, $\mathrm{DF} = 13$, $p > 0.000$) and for presentation form ($F = 9.58$, $\mathrm{DF} = 6$, $p > 0.000$). Hence, both tasks and presentation forms varied in perceived complexity. In order to check whether eventual differences between the tasks or the presentation forms would influence the effects of presentation form on performance, an interaction analysis between task and presentation form was calculated. No interaction effect was found ($F = 0.95$, $\mathrm{DF} = 78$, $p = 0.593$), hence, the disparity of task and presentation form had no systematically combined effect on the performance.

4.3. General effects of knowledge

We did not calculate a main effect of knowledge because the three knowledge categories were not directly comparable. However, a qualitative analysis of the data shows that the verbal facts knowledge (mean = 1.6) scored higher than the visual knowledge (mean = 1.38), mean rank = 4.04 and 2.5, respectively. The possible range of scores for visual knowledge and verbal facts knowledge were between 0 and 2 and for inference knowledge between 0 and 100%. (The average score for inference knowledge was 49.43%.) Hence, an estimated comparison of correct scores between the three knowledge measures can be ranked bottom up as: visual knowledge below 50%, inference knowledge about 50% and verbal facts knowledge above 50%.

4.4. Effects of presentation forms on knowledge

The Friedman test was used to analyse the effects of the presentation forms for the verbal facts- and visual knowledge learning performance data. The Wilcoxon rank test was used to map the exact differences between the presentation forms. The inference knowledge question could be analysed with an ANOVA, and the comparison of the different presentation forms was computed with Fischer's PLSD. Table 2 shows the medium ranks and the mean values for the dependent measures.

Effects for visual knowledge: significant differences between the presentation forms (with a picture) were found ($\chi^2 = 10.03$, $p = 0.018$). Table 3 shows the results of the Wilcoxon test analysing the effects of the different possible combinations.

Effects for verbal facts knowledge: significant differences between the presentation forms were found ($\chi^2 = 18.902$, $p = 0.004$). Table 3 shows the results of the Wilcoxon test analysing the effects of the different possible combinations.

Effects for inference knowledge: Significant differences between the presentation forms were found ($DF = 6$, $F = 7.77$, $p = 0.000$). Table 3 shows the results of Fischer's PLSD analysing the effects of the different possible combinations.

4.5. Effects of the presentation forms on preferences

There were significant differences in preference between the presentation forms (Friedman: $\chi^2 = 26.756$, $p = 0.000$). The differences between the presentation forms were tested with Wilcoxon rank correlation. Table 2 shows the mean preference values for each presentation form and Table 3 shows the results of the comparison.

4.6. Effect of the control variables

The Benton test revealed normal results in respect to visual memory skills. Hence, none of the participants

Table 3
Statistical comparisons of the presentation forms by the dependent measures

Dependent measure	Comparisons	Z	p
Visual knowledge	P > T + V + P	−2.214	0.027
	V + P > T + P	−2.319	0.021
	V + P > T + V + P	−2.561	0.010
Verbal facts knowledge	V + P > V	−2.147	0.032
	T + V > V	−2.264	0.024
	V + P > P	−3.349	0.000
	T > P	−2.502	0.012
	T + V > P	−3.182	0.002
	T + V + P > P	−2.298	0.022
	T + V > T + P	−2.523	0.012
		Mean difference	p
Inference knowledge	V + P > V	−32.88	0.000
	T + V > V	−24.25	0.000
	T + P > V	−14.65	0.031
	T + V + P > V	−22.92	0.000
	V + P > P	−32.47	0.000
	T + V > P	−23.83	0.000
	T + P > P	−14.23	0.036
	T + V + P > P	−22.50	0.001
	V + P > T	30.88	0.000
	V + P > T + P	18.24	0.007
	T + V > T	−20.91	0.001
	T + V + P > T	−2.298	0.002
Preference	V > T + V	−2.133	0.033
	V > T + V + P	−2.553	0.012
	T > P	−2.628	0.008
	T > V + P	−2.952	0.003
	T > T + V	−2.471	0.014
	T > T + V + P	−2.624	0.009
	T + P > T + V	−2.225	0.026
	T + P > T + V + P	−2.274	0.023

Table 2
Medium and mean values for the dependent measures

Presentation forms	Medium rank		Inference	Preference
	Facts	Visual	Mean %	Mean
Text (T)	4.119	—	37.563	4.951
Voice (V)	3.774	—	35.555	4.463
Picture (P)	3.262	2.167	35.976	3.435
Voice + picture (V + P)	4.345	2.789	68.444	3.463
Text + voice (T + V)	4.512	—	59.802	3.683
Text + picture (T + P)	3.738	2.357	50.206	4.488
Text + voice + picture (T + V + P)	4.250	2.679	58.476	3.512

Table 4
Correlations between scholastic and learning performance

School subject	Visual			Verbal facts			Inference		
	Rho	*Z*	*p*	Rho	*Z*	*p*	Rho	*Z*	*p*
Math				0.37	2.38	0.02			
German	0.32	2.05	0.04	0.66	4.25	0.000	0.36	2.28	0.02
Local studies				0.49	3.16	0.000	0.46	2.97	0.003
Drawing				0.32	2.07	0.04			

were excluded based on impaired visual memory skills. Furthermore, there were no effects of visual memory skills on any of the dependent variables. In addition, no effects of general memory skills were found.

We performed a correlation analysis between general and visual memory, scholastic performance and the results on learning performance (Spearman rank correlation). Only a positive correlation between scholastic performance and learning performance was found. Pupils with good scholastic performance also performed well in the experiment. Table 4 gives an overview of the results. There was no correlation between scholastic performance and media, which indicates that this result had little effect on the results in general. There was no effect of school performance on the satisfaction to take part in the experiment. In addition, the experiment showed no effect of gender or age on the performance.

5. Discussion

5.1. Main effects

The between group design enabled the presentation of a large variation of tasks, which we regarded as important in order to rule out the fact that participants could influence the performance by talking about the tasks. The analysis showed that the randomization of tasks to the two groups resulted in a minor difference between the two groups in handling the tasks (pause between the tasks). However, this effect did not influence the performance between the two groups. We therefore calculated the results based on the total sample of participants and tasks.

The analysis of the main effects showed that the performance varied significantly between tasks. Hence, the tasks were perceived as disparate. The fact that we ran the analysis employing the means per performance measure over the 14 tasks minimized this effect. In addition, there were no interaction effects between task and presentation form, which shows that the task variations did not influence the effects of the presentation forms. The analysis showed several differences between the presentation forms, which will be reported below. The analysis of the control variables supports the validity of the interpretations of the main results.

5.2. Effects of the presentation forms

The influence of the presentation forms on learning shows some interesting effects also beyond those assumptions stated in the hypotheses. In general, the results support the media principles, but effects are also found that are not covered in these principles. The effects of the presentation forms are not consistent and vary over the knowledge categories.

5.2.1. Visual knowledge

The results for visual knowledge generally support the ''modality principle'' and the ''redundancy principle'', but not clearly the ''multimedia principle'' (Mayer, 2001). As expected, the combination 'voice + picture' resulted in better visual knowledge than 'text + picture'. This multimedia combination enabled the participants to study the information in the picture while being guided by the voice. They could optimally use the short learning time because they did not have to split their attention between reading and visual search. The results also show that the presentation forms 'picture only' as well as 'voice + picture' resulted in better visual knowledge than the triple combination. This is a strong result in respect of CL, i.e. redundancy. The external CL related to the triple presentation distracted the attention away from the pictorial information. In comparison, the participants integrated more visual information when they could study only the picture.

5.2.2. Verbal facts knowledge

The results for verbal facts knowledge support the 'multimedia principle' but only partly the 'modality principle'. The results show that verbal facts knowledge is more easily acquired when the information is presented dually coded with at least one verbal component, either voice or text. Remarkably the combination 'text + picture' was more disturbing for learning verbal facts than the combination 'text and voice'. The dual verbal presentation 'text + voice' resulted in better knowledge than 'voice only', 'picture only' as well as for 'text + picture'. Hence, an eventual split attention effect between reading and listening did not impair the learning. This unexpected effect could have been caused by the fact that the participants were in fourth grade and had just learned to read properly. They may have preferred to hear instead of reading the text themselves. Hence, these results may be different with adult participants. Anyway, the results suggest that negative effects of split attention may therefore be related to the characteristics of the knowledge category rather than being general. When a picture was presented simultaneously to text, the picture seemed to withdraw the attention from the verbal facts learning—more than the split attention between 'voice + text'. The 'multimedia principle' was only partially supported. The results showed that the presentation of 'text only' was better than 'picture only'. Not surprisingly, the results show that the presentation of a picture is not relevant for learning verbal facts: The presentation forms 'text only', 'text + voice', 'voice + picture' as well as 'voice + picture + text' resulted in better learning than when a picture was presented alone. Hence, for the learning of verbal facts there is a strong connection between the modality of the presentation form and the modality of the resulting knowledge. Consequently, this also indicates that the test of knowledge should be related to the presentation mode, i.e. it would be unjust to expect exact verbal facts knowledge from strongly pictorially presentations.

5.2.3. *Inference knowledge*

The results for inference knowledge show that it is crucial that the presentation offers rich information. Almost regardless what media-combination, results with dual and triple combinations proved better than results reflecting mono presentations (voice, text or picture). The results support the 'multimedia principle' and the 'modality principle': the combinations 'voice + picture' proved better than 'text + picture' and 'text only'. The results only partly support the 'redundancy principle'. This could indicate that the intrinsic CL, i.e. task complexity, was lower than expected, and that free capacity could be used to compensate the high external CL. However, this seems not to explain the result because we then would have expected similar effects for verbal facts knowledge, which was not the case. Further, we expected that mono picture presentations would imply a disadvantage to the other mono presentations with text and voice for inference knowledge. This seems not to be the case, because all the mono presentation forms resulted in poorer inference knowledge than the multimedia presentations.

5.2.4. *Effects of the presentation forms on preference*

The preference for the presentation forms show that in most cases the participants preferred the presentation forms that objectively seen did not optimally support learning, i.e. the mono presentations or the 'text + picture' forms. This indicates that the subjective threshold for external CL was low. The presentation form 'picture only' was the least preferred. This is not surprising because verbal facts and inference knowledge could not be learned with germane CL in that condition. High CL was imposed upon the pupils when they had to actively recode the pictorial information into semantic entities that could be integrated into a logical model. Generally, the results on preference show that CL has a largely subjective component, which, although it seems to have no immediate effect on the learning performance, may influence motivation towards the learning context when the duration of the learning session increases.

5.3. *Evaluation of the hypotheses*

Hypothesis one was not clearly supported in this experiment. The results for 'text + voice' varies over the knowledge categories: For verbal knowledge 'text + voice' combinations had a better impact on learning than 'text + picture', and for inference knowledge 'text + voice' supported learning better than the mono presentation forms. Hence, the modality specific split attention assumption, i.e. double load on the verbal channel, may not be relevant for all kinds of learning. Depending on the knowledge category, our results show that 'text + voice' even can support learning. This is an interesting result because it suggests that the split attention assumption should be moderated and related to the characteristics of the knowledge category.

Hypothesis two was supported. External CL on the visual processing channel is detrimental for learning. Visual split attention was particularly detrimental for facts knowledge and for visual knowledge.

Hypothesis three was supported, but not exclusively over all the knowledge categories. A clear benefit of 'picture + voice' combinations was only found for visual knowledge and partly for inference knowledge. However, the 'picture + text' combinations did not result in higher learning scores than the 'picture + voice' combinations for any of the knowledge categories. Hence, our results deliver support for the 'modality principle'.

Hypothesis four was clearly supported for inference knowledge. Mono presentations with picture or voice were particular detrimental for verbal facts knowledge. The preference measure shows an interesting opposite effect, indicating that the mono presentation forms were the only conditions offering optimal CL. This result supports the 'multimedia principle'.

The results only show partial support for *Hypothesis five*. There were signs of CL trough the triple media combination (the 'redundancy principle'), but this presentation form was not exclusively detrimental. High external CL showed to be most critical for learning visual knowledge. For the other knowledge categories the disadvantage of the triple media combinations was less obvious. Therefore, also the effect of very high external CL may depend on the knowledge measures. The subjective measure, however, clearly shows that the triple media combinations invoke too high external CL.

6. Conclusions

This experiment has brought interesting insight about media effects, suggesting that the effect of (multi-) media should be evaluated relative to the learning goals and the respective measures. People do not either learn or not learn, rather various kinds of knowledge results from different combinations of verbal and visual presentation forms. Hence, our results did not generally contradict Mayer's principles, but rather showed that the extent to which certain media combinations support learning is a mater of how well the media (combinations) fit with the learning goal.

The CL theory offers a sound frame for understanding the effects of (multi) media presentations. However, it is necessary to relate CL to a more complete model or theory of multimedia learning in order to fully understand multimedia learning. We did not aim at an exact matching of our results with a theoretical model. This is, however, a necessary further step in order to completely understand the processing of verbal and visual media. The experiment has shown that external CL through split attention may not be generally detrimental but depends on the learning task and measures. The results showed that verbal processing is less affected from split attention ('text + voice') than visual split attention ('text + picture'). Hence, open theoretical questions with relevance to this study continue to address how verbal and visual presentations are processed and represented in our cognitive system and how to explain the differences in CL from split attention on the verbal and visual processing channel.

The conclusions on the level of knowledge testing and research are related. On the one hand, experimental research in this domain should apply a more differentiated knowledge concept than the case often is today. When knowledge representation is closely linked to information presentation, it should have consequences for the design of experiments in this domain. On the other hand, our results also have general implications for how to test knowledge in schools and universities. Knowledge resulting from a multimedia learning process corresponds to the modality of the presentation. Hence, our knowledge representations resemble the real perceptual experiences. The test of knowledge should therefore correspond to the presentation form. Moreover, when the learning goal has a predefined modality (visual or verbal) the presentation should support the formation of such representations. Further research supporting this suggestion should have impact on the testing of scholastic performance in general.

References

Benton, A.L., Sivan, A.B., Hamsher, K., Varney, N.R., Spreen, O., 1994. Contributions of Neuropsychological Assessment: A Clinical Manual 2nd, Oxford.

Engelkamp, J., 1991. Das menschliche Gedächtnis (2. Auflage). Verlag für Psychologie, Göttingen.

Guttormsen Schär, S., 2006. Multimedia Didactics. Studies on the Learning Effects of Dynamic Visual Media, Based on Analyses of Cognitive Models and Learning Content Characteristics. Shaker Verlag, Aachen.

Guttormsen Schär, S., Messerli, S., Schluep, S., 2004. Information categories for learning-object metadata. An empirical study on the practical validity of theoretically founded information categories. In: The Proceedings of Ed-Media 2004, Lugano, Switzerland.

Hamburg-Wechsler Intelligenztest für Kinder, Revision 1983, (HAWIK-R) by Tewes & Titze.

Hasebrook, J., 1995. Multimedia-Psychologie: eine neue Perspektive menschlicher Kommunikation. Spektrum Akademischer Verlag, Heidelberg.

Hegarty, M., Narayanan, N.H., Freitas, P., 2002. Understanding machines from multimedia and hypermedia presentations. In: Otero, J., Graesser, A.C., Leon, J. (Eds.), The Psychology of Science Text Comprehension. Erlbaum, Mahwah, NJ, pp. 357–384.

Kirschner, P.A., 2002. Cognitive load theory: implications of cognitive load theory on the design of learning. Learning and Instruction 12, 1–10.

Kirschner, D., Whitson, J. (Eds.), 1997. Situated Cognition: Social, Semiotic and Psychological Perspectives. Lawrence Erlbaum Associates, Mahwah, NJ.

Large, A., 1996. Computer animation in an instructional environment. LISR 18, 3–23.

Lave, J., Wenger, E., 1991. Situated Learning. Cambridge, New York.

Lewalter, D., 2003. Cognitive strategies for learning from static and dynamic visuals. Learning and Instruction 13, 177–189.

Lowe, R.K., 2003. Animation and learning: selective processing of information in dynamic graphics. Learning and Instruction 13, 157–176.

Paivio, A., 1986. Dual coding theory. In: Paivio, A. (Ed.), Mental Representations. A dual Coding Approach. Oxford University Press, Oxford, pp. 53–83.

Pollock, E., Chandler, P., Sweller, J., 2002. Assimilating complex information. Learning and Instruction 12, 61–86.

Mayer, R.E., 2001. Multimedia Learning. Cambridge University Press, Cambridge.

Mayer, R.E., 2003. The promise of multimedia learning: using the same instructional design methods across different media. Learning and Instruction 13, 125–139.

Mayer, R.E., Moreno, R., 1998. A split attention affect in multimedia learning: evidence for dual processing systems in working memory. Journal of Educational Psychology 90, 312–320.

Mayer, R.E., Moreno, R., 2002. Animations as an aid to multimedia learning. Educational Psychology Review 14 (1), 887–899.

Mayer, R.E., Hegarty, M., Mayer, S., Campbell, J., 2005. When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. Journal of Experimental Psychology: Applied 11 (4), 256–265.

McLellan, H., 1995. Situated Learning Perspectives. Educational Technology Publications, Englewood Cliffs, NJ.

Merrill, M.D., 1983. Component Display Theory. In Instructional-Design Theories and Models: An overview of their Current Status. Lawrence Erlbaum Associates, Hillsdale.

Scaife, M., Rogers, Y., 1996. External cognition: how do graphical representations work? International Journal of Human Computer Studies 45, 185–213.

Schnotz, W., Bannert, M., 2003. Construction and inferences in learning from multiple representation. Learning and Instruction 13, 141–156.

Sweller, J., 1999. Instructional Design (Vol. 43). Acer Press.

Turban, E., Aronson, J., 1988. Decision Support Systems and Intelligent Systems. Prentice Hall Inc., Upper Saddle River, NJ.

Tversky, B., Morrison, J.B., Betrancourt, M., 2002. Animation: can it facilitate? International Journal of Human–Computer Studies 57, 247–262.

Wiener, B.J., 1971. Statistical Principles in Experimental Design. McGraw-Hill, New York, NY.