Computers & Education 100 (2016) 18–31

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

One for all?! Simultaneous examination of load-inducing factors for advancing media-related instructional research

Maria Wirzberger*, Maik Beege, Sascha Schneider, Steve Nebel, Günter Daniel Rey

E-Learning and New Media, Institute for Media Research, Faculty of Humanities, Technische Universität Chemnitz, Germany

A R T I C L E I N F O

Article history:
Received 19 February 2016
Received in revised form 15 April 2016
Accepted 20 April 2016
Available online 22 April 2016

Keywords:
Cognitive load theory
Working memory updating
Task complexity
Split-attention effect
Schemata

A B S T R A C T

In multimedia learning settings, limitations in learners' mental resource capacities need to be considered to avoid impairing effects on learning performance. Based on the prominent and often quoted Cognitive Load Theory, this study investigates the potential of a single experimental approach to provide simultaneous and separate measures for the postulated load-inducing factors. Applying a basal letter-learning task related to the process of working memory updating, intrinsic cognitive load (by varying task complexity), extraneous cognitive load (via inducing split-attention demands) and germane cognitive load (by varying the presence of schemata) were manipulated within a 3 × 2 × 2-factorial full repeated-measures design. The performance of a student sample ($N = 96$) was inspected regarding reaction times and errors in updating and recall steps. Approaching the results with linear mixed models, the effect of complexity gained substantial strength, whereas the other factors received at least partial significant support. Additionally, interactions between two or all load-inducing factors occurred. Despite various open questions, the study comprises a promising step for the empirical investigation of existing construction yards in cognitive load research.

1. Introduction

Learning demands a variety of cognitive processes related to information capture, storage, and retrieval that request learners' mental resources. It involves in particular those associated with memory structures, entailing the challenge to keep track of changing contents in working memory, and their correct and stable representation in long-term memory. Particularly within multimedia learning settings, learners' limited mental resource capacity has to be taken into account to avoid impairing overload. Despite their enhanced potential in capturing motivation and engagement, such settings are prone to overly claim mental resources due to the multimodal, interactive and often distributed presentation of subjects. To be able to handle this opportunities in a balanced and constructive manner, the necessity of a closer investigation of factors and effects related to mental resource demand arises. A prominent and influential theory providing advices for the conducive design of media-transmitted instructional content is the Cognitive Load Theory (CLT). It was introduced in the late 1980s by John Sweller (1988) and emerged a well-known and extensively used approach. Nevertheless, several construction yards exist

---

* Corresponding author. E-Learning and New Media, Institute for Media Research, Faculty of Humanities, Technische Universität Chemnitz, Straße der Nationen 12, 09111 Chemnitz, Germany.

E-mail address: maria.wirzberger@phil.tu-chemnitz.de (M. Wirzberger).

http://dx.doi.org/10.1016/j.compedu.2016.04.010
0360-1315/

within this framework, above all issues of a valid and reliable empirical assessment of the theoretically postulated building blocks and assumptions regarding their coaction. The current research accepts the emerging challenges and contributes to their clarification, to be able to derive more detailed predictions on underlying learner cognition within a next step.

1.1. *Cognitive Load Theory*

Amongst its basic assumptions, the CLT postulates a practically unlimited storage capacity of long-term memory, the mental representation and organization of knowledge via schemata, and a limitation of working memory in terms of duration and capacity. Additionally, a separation of the overall cognitive load (CL) construct into different facets related to distinct aspects within a learning setting has been assumed during the last decades (Sweller, Ayres, & Kalyuga, 2011). While intrinsic cognitive load (ICL) should result from the complexity of the used learning material (referred to as element interactivity) and takes into account a learner's previous knowledge, extraneous cognitive load (ECL) arises from the instruction itself, for instance by containing interesting but irrelevant content or demanding learners to spread their attention across different sources of information. Relevant processes of schema acquisition and automation, which represent crucial aspects while learning certain contents, are assigned to germane cognitive load (GCL). Such CL types should operate additively on the available amount of cognitive resources, implying an increase in relevant processing just in the case irrelevant processing decreases. However, recent research queries the assumption of additivity (Brünken, Plass, & Moreno, 2010; Park, 2010) as well as the separability of load facets (Kalyuga, 2011; de Jong, 2010), not least due to the lack of satisfying means of measurement related to the described CL facets (Brünken, Seufert, & Paas, 2010). Yet another step forward, Sweller (2010) aimed at reformulating the three-factorial framework by attributing germane resources to handle content relevant to achieving a defined learning outcome (ICL) and extraneous resources to deal with irrelevant situational characteristics (ECL). Such dual framework would take into account the fact of certain load to be beneficial for learning, but on the other hand presumes each learner's motivation to spend all available resources to the process of learning (Kalyuga, 2011).

So far, a sophisticated approach to empirically test the assumption of three additively operating load factors was applied by Park (2010) within a series of learning experiments that varied either ECL, GCL or both. ICL was kept at a constant level because it was considered to be rather stable and hardly influenceable by instructional design. Attempting to explain her results, Park (2010) states that the emerging pattern of non-significant main effects and significant interactions strongly challenges the additive contribution of the postulated load-inducing factors. Nevertheless, the chosen approach faces certain limitations. First of all, none of the experiments comprised a variation of ICL. However, a comprehensive examination of separate and additive influences should address and manipulate all facets within the same framework. In addition, dependent variables comprised subjectively rated amounts of cognitive load, and scales on learning success with varying amounts of retention, transfer, and problem comprehension for each experiment. Objective measures related to defined behavioral outcomes might be an alternative to facilitate more universal predictions.

1.2. *Task complexity and ICL*

Advancing the matter of task complexity, associated with the facet of ICL, Sweller and Chandler (1994) postulate that beyond the amount of information the resource demand induced by a learning task arises from related information that has to be processed simultaneously. In doing so, they outline the crucial role of interactivity between elements of a learning task, whereat elements could be symbols, concepts, procedures or other types of task inherent units (Chen, Kalyuga, & Sweller, 2016). These are measurable a priori for instance by counting the number of separable but interdependent subtasks. Subtasks comprise defined cognitive acts that rely on learners' cognitive resources, and are demanded to various extents for differences in existing knowledge on the presented content. A felicitous implementation of a priori estimates of task complexity was introduced by Beckmann (2010). He used an abstract reasoning task with geometric symbols, and increased the level of complexity by varying the number of dimensions presented items differed on, ranging from two (shape and color) to four (shape and color of inner as well as outer components). Such controlled approach allows to give concise predictions about cognitive acts that have to be performed while solving the task, to quantify the extent of complexity in a reliable manner. Besides a significantly worse performance with increasing complexity ($\eta_p^2 = .37$) the obtained results reveal a better performance without the requirement to store results of individual subtasks ($\eta_p^2 = .47$). The arising predictability of performance outcomes supported the chosen approach. Moreover, Beckmann (2010) emphasized that apart from task-related characteristics those related to the respective situational context contribute to overall task complexity as well.

1.3. *Split-attention effect and ECL*

The situational aspect of instructional design generally relates to the facet of ECL, resulting in design principles to avoid distracting overload. An often studied phenomenon in this context is the split-attention effect (Chandler & Sweller, 1991, 1992; Owens & Sweller 2008), occurring in learning with various sources of information. Given that each source of information matters for understanding the learning material, learning outcomes improve when different sources of information are presented spatially integrated rather than in a separated format. An explanation assumes that in the latter case information must be maintained in working memory, while searching for elements within distributed but interconnected sources (Sweller et al., 2011). Such additional demands potentially reduce the capacity available for relevant learner involvement, and are prone to decrease learning performance. By contrast, if instructional sources of information are presented in an integrated format, learners are less demanded to split their attention, and a higher amount of working memory capacity can be dedicated to relevant processes of learning. Similarly, the spatial contiguity principle, based on the Cognitive Theory of Multimedia Learning (CTML; Mayer & Moreno, 1998; Mayer, 2014) postulates that various sources of information should be presented close to each other to foster learning. In his meta-analysis supporting the split-attention effect, Ginns (2006) furthermore outlined that harms and benefits of spatially split vs. integrated information depend on the complexity of certain learning materials, determined by the extent of element interactivity. In the case of high element interactivity and/or no or low prior knowledge, integration can be characterized as efficient and effective regarding instructional quality, obvious due to rather strong effects ($d = 0.78$) according to conventions on effect sizes stated by Cohen (1988). On the other hand, if element interactivity is low, even split information has only a weak effect ($d = 0.28$). Such results align to the element interactivity effect, stating that design effects affect performance only under high amounts of interrelated elements, whereas low amounts can compensate for inappropriate and demanding instructional designs (Sweller & Chandler, 1994).

1.4. Schemata and GCL

Schemata are characterized as organized patterns of knowledge (Kalyuga, 2010; Sweller & Chandler, 1994), and constitute crucial elements when approaching the facet of GCL. If learners have enough resources available, they are able to build up relations within the learning material. Such process was described as coherence formation in corresponding research (e.g. Park, 2010; Seufert, 2003; Seufert & Brünken, 2006). According to Schnotz and Kürschner (2007), activities going beyond simple task performance comprise relevant aspects in this context. They explicitly named the process of intentionally searching for patterns within the presented learning material, on the purpose to abstract cognitive schemata and create semantic macrostructures. A task qualified to elicit such processes can hold long-term effects on performance, since once generated schemata are stored in long-term memory, and become parts of learners' previous knowledge. On this account, they codetermine resource demands throughout the subsequent learning process (Kalyuga, 2010).

1.5. Working memory

An important source of constraints in information processing exists as a result of working memory resource limitation, both in terms of duration and capacity (Wickens, Hollands, Banbury, & Parasuraman, 2013). While the first aspect refers to the fact of information decay in working memory after a certain time, the matter of capacity indicates that just a defined amount of information can be stored there at the same time. According to Miller (1956) this should reside between five and nine items, although more recent research proposes a smaller number of about four elements (Cowan, 2010; Cowan, Morey, & Chen, 2007). Within the theoretical framework of the CLT, working memory plays a crucial role when explaining how learning tasks rely on learners' cognitive resources. Besides that, the theory holds connections to the concept of long-term memory as well, since learning involves the development of schemata that are stored on a longer run. In this regard, Schweppe and Rummer (2014) describe working memory as activated part of long-term memory (Cowan, 1999), and incorporate the aspect of attention in terms of focused resources.

Since learning involves dealing with altering information, the construct of working memory updating (WMU) bears high relevance, as changing working memory content should be represented correctly over time. It comprises three constituting features that independently contribute to updating performance (Ecker, Lewandowsky, Oberauer, & Chee, 2010). While retrieval consists of extracting relevant information from memory, transformation can be identified as adjusting this information according to situational changes. Finally, substitution results in replacing the previous informational state by the current one, entailing an updated content representation in working memory. All described components have been confirmed experimentally and were applied in WMU tasks to various extents. After several steps of updating, participants are usually required to recall the final state of the previously presented information entities. Such recall requires storage processes on a longer term, comprising an additional benefit when inspecting task-related performance. Additionally, this measure aligns well to the crucial role of limited working memory capacity in the CLT.

1.6. The present study

The current study investigates the potential of a single experimental approach to provide simultaneous and separate measures for the three-factorial framework of cognitive load facets. Such allows to manipulate each facet in a selective, controllable way, and directly relates behavioral outcomes like task-related timing or errors to the process of learning. In this vein, it provides a benefit compared to collecting indirect subjective responses via questionnaires or applying time and resource consuming physiological measures that often lack sensitivity and diagnosticity (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Verwey & Veltman, 1996). Park (2010) already recommended the inclusion of aptitude variables in research concerned with the CLT, like working memory capacity or specific memory skills, whereas Brünken, Plass et al. (2010) suggest to integrate new paradigms from basic research into CL measurement. Both support the use of a task related to elementary working memory research, for instance a task involving processes of WMU. However, this could be regarded as learning task as well, since people aim to remember defined content and retrieve it later, similar to retention performed in explicit learning tasks.

1.7. Hypotheses

Approaching the load-inducing factors individually, for the facet of ICL Sweller and Chandler (1994) outlined the crucial role of interrelation between task elements when rating task complexity. Such can be evaluated a priori by estimating the number of related subtasks performed within a task (Beckmann, 2010).

Hypothesis 1. *A higher amount of task complexity increases demands on learners' cognitive resources and fosters a substantial decrease in performance.*

Furthermore, referring to the results reported by Chandler and Sweller (1991, 1992) concerning the facet of ECL, in the case that information relevant to a certain learning task is spatially distributed across different sources, learners have to spend more cognitive resources to cope with the task.

Hypothesis 2. *The necessity to spatially split attention puts additional demands on learners' cognitive resources and results in decreased performance.*

The facet of GCL postulates that successful learning fosters the development of cognitive schemata from obtained knowledge (Kalyuga, 2010). The opportunity to rely on such previously developed schemata while performing a certain task is assumed to relieve learners' cognitive resources (Schweppe & Rummer, 2014).

Hypothesis 3. *Due to the presence of schemata, learners' cognitive resources are less demanded and facilitate an increased performance.*

As postulated by Sweller et al. (2011), the outlined facets of CL are assumed to demand cognitive resources in a strictly additive manner. In consequence, arising effects should show a pattern of independence among themselves, whereas substantial interrelations would query that theory-based assumption.

Hypothesis 4. *No interactions between the manipulated facets are postulated.*

2. Methods

2.1. Participants

A total of 96 university students ($M_{\text{age}} = 24.35$ years, $SD_{\text{age}} = 4.81$, 76 female) participated in the study. The sample split up into various disciplinary backgrounds, comprising Psychology (26%), Education (21%), Communication Sciences (29%), and other social and technical subjects (24%). Regarding language skills, participants were either native German speakers (97%) or actively spoke the language for at least 12 years. For compensation, they received a financial allowance of 5 € or course credits according to their curriculum.

2.2. Design

Hypotheses were tested with a 3 × 2 × 2-factorial, multivariate within-subjects design including complexity (low vs. medium vs. high), split attention (with vs. without) and schema presence (with vs. without) as independent variables. Reaction times and errors in update and recall trials comprised the dependent variables. Since individual differences in the ability to focus attention exert influence within memory tasks, concentrated attention was recorded prior to completing the main task. Moreover, perceived mental effort, task difficulty, and clarity of instruction were inquired to ensure an adequate level of complexity, and participants' involvement and understanding of the task. Due to the arising hierarchical design (multiple observations nested within each participant), a linear mixed model approach, often referred to as hierarchical linear, multilevel, random effects or mixed effects modeling (Garson, 2013), was chosen to inspect the hypothesized relationships in a more adequate manner.

2.2.1. Independent variables

Independent variables were addressed according to the theoretical descriptions of the CL facets within a WMU task, adapted from Ecker et al. (2010). The task consisted of 24 trials that required updating and memorizing an initially presented letter set during six steps of alphabetic transformations. Task complexity was manipulated by varying the number of letters displayed at the outset of a trial. It comprised three levels of difficulty, appearing with equal frequency during the task, that is two, three or four letters to remember and transform within a trial. The decision for such definition of levels was based upon Beckmann (2010), using items with increasing dimensionality (two up to four) to achieve different levels of task complexity, and in this vein set up diverse levels of ICL. On the purpose to manipulate ECL, the horizontal spatial distance between the displayed elements was scaled up in half of the trials, to induce the demand to split up attentional resources. Finally, the facet of GCL was addressed by the opportunity to build up task-related schemata on presented letter sets during a preceding practice sequence. Within the test sequence, those letter sets were fully or partly repeated in half of the trials, enabling participants to rely on previously acquired patterns of knowledge.

2.2.2. Dependent variables

Regarding dependent variables, reaction times and correctness of task responses were recorded during the WMU task. Although Ecker et al. (2010) only focused on letter updating performance, the final recall of all remembered letters in the end of a trial was taken into account as well in this experiment. Such decision was made for the assumption of distinct cognitive features underlying update and recall processes. Whereas updating requires the initially outlined transformation steps, recall represents more static aspects like duration and capacity of storing information. Both are considered as highly relevant to the concept of working memory constraints as core assumption of the CLT.

2.2.3. Aptitude and control variables

The standardized psychological attention and concentration inventory d2-R (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010) addressed participants' ability to concentrate attention on a certain task. Finally, three questions dealing with the aspects of perceived mental effort, task difficulty and clarity of instruction were used. The first question on perceived mental effort was directly adapted from Paas et al. (2003), who often combine such a rating with an estimation of task difficulty (Brünken Seufert, & Paas, 2010). On this account, the second question referred to the perceived task difficulty, whereas the last question covered the perceived clarity of instruction.

2.3. Material

2.3.1. WMU task

Stimuli were presented on desktop computers with a screen size of 24″, a screen resolution of 96 dpi, a display resolution of at least 1680 × 1050 px and a video refresh rate of 75 Hz. The task was implemented in PXLab (Irtel, 2007) with a timing

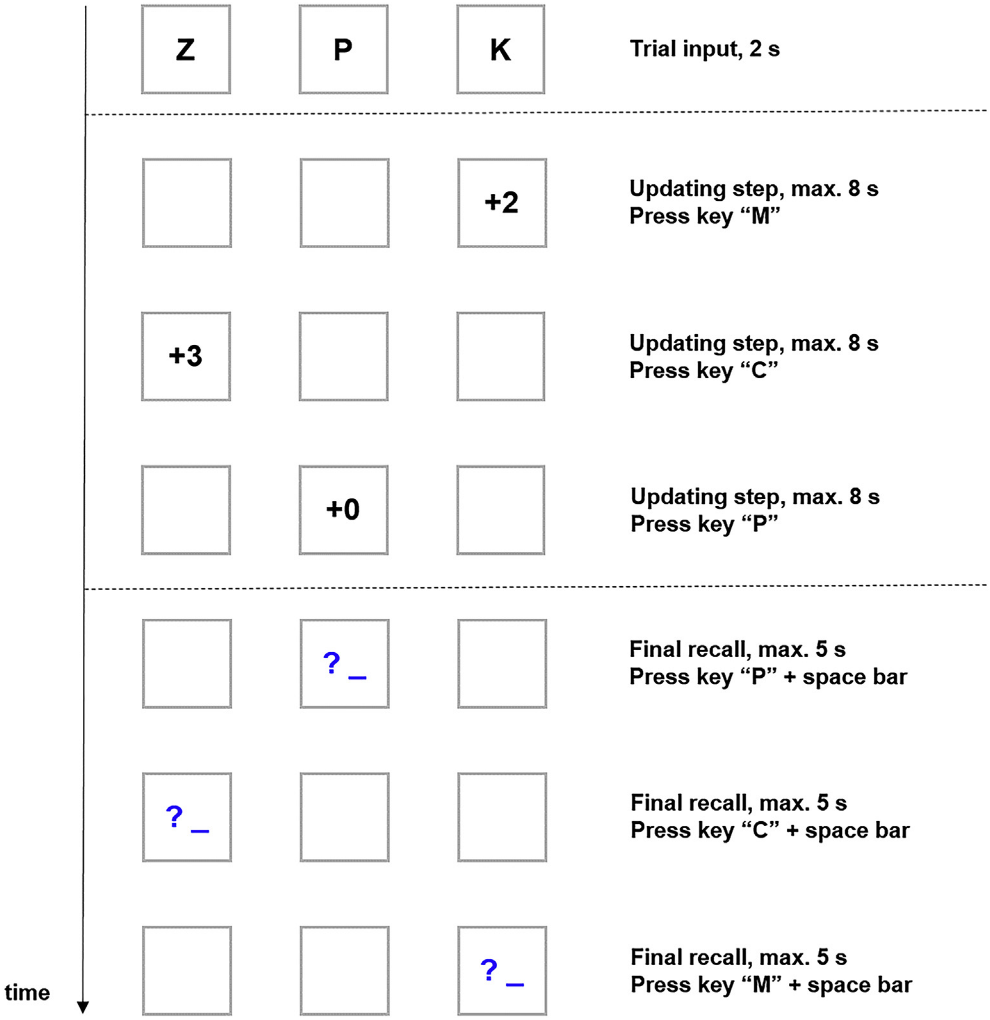

Fig. 1. Sample practice trial sequence for the working memory updating task.

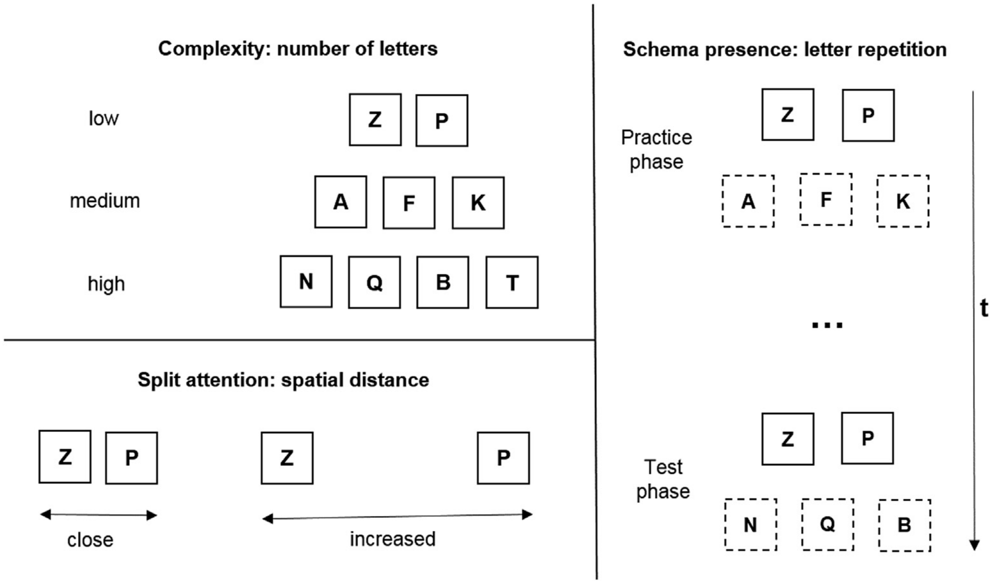

Fig. 2. Experimental manipulations of complexity, split-attention and schema presence. Boxes with dashed lines indicate the lack of repetition in letter sets.

precision better than 1 ms. After performing a written instruction including a detailed example, a set of six practice trials followed. As illustrated in Fig. 1, initially two, three or four framed letters from the Latin alphabet appeared in row for 2 s. In line with Ecker et al. (2010), between presented input letters a minimal alphabetic distance of five was chosen. Letters vanished after the indicated time span, and an updating instruction referring to one of the letters was displayed. Participants had to increment the indicated letter by zero, one, two or three positions in the alphabet, and type in the result within a time frame of 8 s, since Ecker et al. (2010) reported mean deadlines of 7.94 s for transformation steps. In line with their work, no visual feedback occurred after typing in the solution, and a lack of response within the time frame was logged as error. To keep the practice sequence short and simple, after a reduced set of three updating steps, the final result of transformations was queried, signaled by blue question marks appearing one by one in each frame. Participants had to type in the indicated letter, received visual feedback on their input, and had to log in their answer by pressing the space bar. Within the final recall period, responses had to be provided within a time frame of 5 s. After completing one trial, the following one started after 2.5 s. These time spans also align to Ecker et al. (2010).

The 24 test trials entailed exactly the same procedure, except the inclusion of six instead of three update steps,[1] and the presentation of each letter set with either close or distant spatial proximity between letters. For close spatial proximity, letter distance resided upon about 80 px for all conditions, whereas in the case of distant spatial proximity it depended on the number of letters, amounting to about 1200 px (two letters), 600 px (three letters) or 400 px (four letters). Fig. 2 provides an impression on the experimental manipulation of the independent variables. After completing all test trials, participants received feedback on the percent of correct responses within the test trials, computed as joint value of update and recall responses. By contrast to Ecker et al. (2010), letters as well as updating instructions appeared in a fixed sequence. This ensured that all participants had to deal with a comparable difficulty of the task.

2.3.2. d2-R

The d2-R (Brickenkamp et al., 2010) assessed the individual level of concentrated attention by demanding to focus on a set of defined target objects while neglecting the presented distractors. Participants received the instruction to cross each small Latin letter “d” accompanied by exactly two dashes, located either above, below or above and below the letter, but not cross a letter “d” with less or more than two dashes or a letter “p” regardless of the number of accompanying dashes. Their task then comprised to complete a test sheet entailing a set of 789 characters grouped in 14 lines with 57 characters each. After a limited time span of 20 s for one line, participants received an experimenter command and proceeded with the following line until they completed the test sheet.

2.3.3. Questions on mental effort, task difficulty and clarity of instruction

Perceived mental effort, task difficulty, and clarity of instruction regarding the working memory updating task were assessed with three individual questions, asking participants to rate each aspect on a nine-point Likert scale from “very, very

---

[1] Such number aligns to the original experimental procedure reported by Ecker et al. (2010).

low" to "very, very high". As already mentioned, the question on mental effort was directly adapted from Paas et al. (2003), whereas the questions on task difficulty and clarity of instruction were self-developed.

2.4. Procedure

The experiment was conducted in a separate learning laboratory, equipped with four desktop computers arranged in a square. Within a testing session, one up to four students participated. They were welcomed, signed the consent form, and filled a questionnaire on demographic aspects. Regarding the d2-R (Brickenkamp et al., 2010), the experimenter first provided instructions according to the test manual, and then participants completed the test. The following WMU task was again preceded by detailed information on how to conduct the task, before participants worked through the practice and test trials at their own pace. Finally, they answered the questions on perceived mental effort, task difficulty, and clarity of instruction regarding the WMU task, received their allowance, were thanked and approved. Experimental sessions lasted about 35–45 min, depending on how fast participants proceeded within the WMU task. Participants also completed a short memory game at the outset of the session that was not integral part of the research focus and thus is not reported in this manuscript.

2.5. Scoring

2.5.1. Dependent variables

Within the WMU task, each key press generated a reaction time value and a response code indicating whether the response had been correct or erroneous. Update and recall steps were evaluated separately by aggregating the respective data points, since blocking is suitable for increasing reliability of constructs, and makes designs more powerful (Stevens, 2009). Reaction times for each trial were calculated via averaging values from the six update steps in the case of update performance ($\text{RT}_{\text{update}}$), or via averaging the two to four observations within the final recall step ($\text{RT}_{\text{recall}}$). An analogous computation was performed for errors ($\text{Errors}_{\text{recall}}$, $\text{Errors}_{\text{update}}$), but sums instead of means were used in this instance. The final error score further took into account the amount of potential responses within a trial, two up to four for $\text{Errors}_{\text{recall}}$ and six for $\text{Errors}_{\text{update}}$. On this account, values between zero and one resulted, indicating the actual amount of errors relative to the possible amount of errors. For $\text{Errors}_{\text{recall}}$ as well as $\text{Errors}_{\text{update}}$ inherited errors were not regarded as actual mistakes but as a result of successful memory performance. In consequence, for each trial their respective amount was subtracted from the total amount of errors, resulting in a corrected error score.[2]

2.5.2. Aptitude and control variables

The d2-R (Brickenkamp et al., 2010) enables the calculation of the individual level of concentration *(KL)*, defined as difference between marked target objects and errors. Raw sum scores can be transferred into standard values afterwards. According to Brickenkamp et al. (2010) the score achieves high reliability, with Cronbach's $\alpha = .96$ over all age groups ($N = 4019$), and $r_{tt} = .85$ after ten days. Finally, scores for the questions on perceived mental effort, task difficulty, and clarity of instruction accrued from the respective marking on the nine-point scale. Since they do not form a shared construct, but rather constitute separate aspects, no overall score was calculated.

3. Results

When completing the d2-R, one participant constantly marked a wrong character, indicating he or she had forgotten the instruction. In consequence, this case had to be removed from the subsequent analyses. Within the WMU task, three participants did not press any key during the updating steps, but performed the update transformation just mentally. In this vein, they constantly achieved reaction times at the maximum trial duration of 8000 ms within the update steps. Aligning to Ecker et al. (2010) they were also excluded from analyses.

Separate linear mixed model analyses for all dependent variables were conducted with the *nlme* package in R (Pinheiro & Bates, 2000; R Core Team, 2015). They operated on restricted maximum likelihood (REML) estimation and comprised within-subjects repeated-measures variables on the first level and between-subjects aptitude variables on the second level (Nezlek, Schröder-Abé, & Schütz, 2006). Participant intercept was included as random effect, whereas the predictor variables complexity, split attention, schema presence and concentration score were treated as fixed effects. To control for potential effects of fatigue, a predictor variable monitoring task processing by counting the respective trial (*task sequence*) was included post-hoc as fixed effect as well. In line with the advice on centering and standardizing (Finch, Bolin, & Kelley, 2014; Gelman & Hill, 2007; Luke, 2004), all variables relevant to the analyses were z-standardized beforehand to obtain standardized regression coefficients. Such provides the opportunity to compare predictive values across variables within the same model and across different models. For all dependent variables, models achieved similar fits on Akaike's information criterion

---

[2] Uncorrected error scores were computed and assessed as well. Since corrected and uncorrected error scores were highly correlated ($\text{Errors}_{\text{recall}}$: $r = .75$, $p < .001$; $\text{Errors}_{\text{update}}$: $r = .95$, $p < .001$) and achieved quite similar effect patterns, only corrected error scores are reported due to their enhanced informative value.

($5164.44 < \text{AIC} < 5438.91$), Bayesian information criterion ($5249.85 < \text{BIC} < 5524.32$) and log-likelihood ($-2704.46 < \text{logLik} < -2567.22$). Compared to baseline models including only random intercept, the predictive ability significantly increased by 52% ($\text{RT}_{\text{recall}}$) to 78% ($\text{Errors}_{\text{recall}}$) due to the full models.

3.1. *Main effects*

Analyses displayed in Table 2 revealed continuously significant effects of complexity for $\text{RT}_{\text{recall}}$, $\text{Errors}_{\text{recall}}$, $\text{RT}_{\text{update}}$ and $\text{Errors}_{\text{update}}$. Positive coefficient values indicate a substantial increase of reaction times as well as errors with increasing complexity. Such assumption is supported by descriptive comparisons of different levels of complexity in Table 1 for all dependent variables. In consequence, results strongly support the first hypothesis that postulates a decrease in learning performance with increasing complexity.

In the case of split attention, analyses displayed significant results for $\text{RT}_{\text{recall}}$ and $\text{RT}_{\text{update}}$. Both coefficient values indicate a small increase in time when attention has to be split up, and descriptive results in Table 1 support such assumption. In this manner, at least regarding reaction times the second hypothesis on decreased learning performance when inducing split attention receives support.

Significant results for schema presence showed up in the case of $\text{Errors}_{\text{recall}}$, $\text{RT}_{\text{update}}$ and $\text{Errors}_{\text{update}}$. Due to the negative coefficient values, results point towards slightly decreased learning performance without the presence of schemata. Descriptive results in Table 1 indeed indicate fewer errors in both update and recall steps and longer reaction times for update steps in trials without schemata compared to those including schemata. On this account, the third hypothesis on increased learning performance due to the presence of schemata is confirmed in most cases as well.

3.2. *Interaction effects*

As displayed in Table 2, for $\text{RT}_{\text{recall}}$, a significant two-way interaction between complexity and schema presence occurred. Coefficient values indicate that the presence of schemata held greater influence on reaction times with an increasing level of complexity. Fig. 3 supports the presumption of different impacts of schema presence according to the respective level of complexity. Whereas no differences are indicated under medium complexity, the presence of schemata marginally increases performance under low complexity, but slightly decreases performance under high complexity.

For $\text{RT}_{\text{update}}$, a significant two-way interaction between complexity and split attention occurred. The negative coefficient indicates a decreasing influence of the demand to split attention on reaction time with increasing complexity. Fig. 4 supports such assumption, since participants performed faster without split attention under low complexity, whereas under medium and high complexity differences between conditions with and without the demand to split attention were only marginal.

For $\text{Errors}_{\text{update}}$, analyses revealed a significant three-way interaction between complexity, split attention, and schema presence. According to Fig. 5, with increasing complexity, interactions between split attention and schema presence become more explicit. Whereas under high complexity error rates were lower in trials with split attention with the presence of schemata, a comparable pattern occurred without the presence of schemata under medium complexity. By contrast, differences in terms of schemata were rather small under low complexity and could only be observed in trials with split attention.

Overall, the occurrence of significant interactions between two or all independent variables queries the independence and additivity of the theoretically postulated CL facets. In consequence, when approaching the fourth hypothesis, with exception of $\text{Errors}_{\text{recall}}$ it seems to stay unsupported.

3.3. *Effects of aptitude and control variables*

Analyses of the d2-R revealed a mean concentration score of 107.11 (*SD* = 9.77) amongst participants and results of the linear mixed model analyses outlined in Table 3 indicate effects of the d2-R score on task-related performance. In terms of

Table 1
Descriptive values of dependent variables regarding main effects of independent variables.

		$\text{RT}_{\text{recall}}$ (ms)		$\text{Errors}_{\text{recall}}$		$\text{RT}_{\text{update}}$ (ms)		$\text{Errors}_{\text{update}}$	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Complexity	2 letters	1854.15	915.30	0.21	0.33	3138.95	892.69	0.24	0.26
	3 letters	2342.29	1057.71	0.45	0.35	3619.10	951.55	0.32	0.25
	4 letters	2575.78	1110.14	0.61	0.31	4107.40	1075.58	0.46	0.27
Split attention	with	2286.98	1068.91	0.42	0.36	3678.09	1040.72	0.34	0.27
	without	2227.83	1077.79	0.44	0.37	3565.54	1062.44	0.34	0.27
Schema presence	with	2252.97	1081.68	0.41	0.36	3577.11	1043.59	0.33	0.27
	without	2261.85	1065.77	0.44	0.37	3666.52	1060.72	0.35	0.27

Note. $\text{RT}_{\text{recall}}$ = reaction time during final recall, $\text{Errors}_{\text{recall}}$ = errors during final recall, $\text{RT}_{\text{update}}$ = reaction time during updating steps, $\text{Errors}_{\text{update}}$ = errors during updating steps. Values based on *N* = 92 participants.

Table 2
Standardized beta-coefficients, standard errors, t-values and significance levels of fixed effects in linear mixed model analyses for independent variables.

	$\mathrm{RT}_{\mathrm{recall}}$ (ms)				$\mathrm{Errors}_{\mathrm{recall}}$				$\mathrm{RT}_{\mathrm{update}}$ (ms)				$\mathrm{Errors}_{\mathrm{update}}$			
	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*
Main effects																
Complexity[a]	0.277	0.015	18.113	<.001	0.442	0.016	27.447	<.001	0.377	0.016	24.037	<.001	0.334	0.016	20.315	<.001
Split-attention[a]	0.033	0.015	2.137	.033	−0.028	0.016	−1.712	.087	0.056	0.016	3.580	<.001	0.001	0.016	0.039	.969
Schema presence[a]	<0.001	0.015	0.016	.987	−0.039	0.016	−2.408	.016	−0.040	0.016	−2.563	.010	−0.045	0.016	−2.709	.007
Two-way interactions																
Complexity × Split-attention[a]	0.025	0.015	1.664	.096	−0.001	0.016	−0.067	.947	−0.059	0.016	−3.760	<.001	0.005	0.016	0.272	.786
Complexity × Schema presence[a]	0.032	0.015	2.069	.039	0.003	0.016	0.227	.820	0.023	0.016	1.446	.148	0.016	0.016	0.946	.344
Split-attention × Schema presence[a]	−0.009	0.015	−0.594	.553	−0.001	0.016	−0.047	.963	−0.019	0.016	−1.224	.221	−0.010	0.016	−0.635	.526
Three-way interaction																
Complexity × Split-attention × Schema presence[a]	−0.007	0.015	−0.429	.668	0.009	0.016	0.530	.596	−0.027	0.016	−1.717	.086	−0.041	0.016	−2.468	.014

Note.
[a] $df = 2105$. $\mathrm{RT}_{\mathrm{recall}}$ = reaction time during final recall, $\mathrm{Errors}_{\mathrm{recall}}$ = errors during final recall, $\mathrm{RT}_{\mathrm{update}}$ = reaction time during updating steps, $\mathrm{Errors}_{\mathrm{update}}$ = errors during updating steps. Values based on $N = 2208$ observations of $N = 92$ participants.Variables for split-attention and schema presence binary coded (0 = with, 1 = without), variable for complexity aligns to number of letters (2/3/4).

$\mathrm{Errors}_{\mathrm{recall}}$, $\mathrm{RT}_{\mathrm{update}}$ and $\mathrm{Errors}_{\mathrm{update}}$, participants achieved smaller values with an increasing level of concentration. Additionally, significant interactions with complexity in the case of $\mathrm{RT}_{\mathrm{recall}}$ and $\mathrm{RT}_{\mathrm{update}}$ showed up, pointing towards stronger differences between conditions in the case of high concentration scores. For $\mathrm{Errors}_{\mathrm{recall}}$, the significant interaction of the d2-R score and split-attention indicates higher deviations between trials with and without split-attention if concentration scores were low. The post-hoc inspected control variable on task sequence clearly objected occurring effects of fatigue, but rather pointed towards training effects for $\mathrm{RT}_{\mathrm{recall}}$, $\mathrm{Errors}_{\mathrm{recall}}$ and $\mathrm{RT}_{\mathrm{update}}$, since negative coefficients indicate a faster and less erroneous task performance.

Taking a look at the questions related to mental effort, task difficulty, and clarity of instruction, most participants perceived the WMU task as quite demanding, obvious by rather high mental effort ratings ($M = 7.83$, $SD = 1.09$). Task difficulty was perceived as high either ($M = 7.82$, $SD = 1.02$), however the task seemed to be clear and understandable from the given instruction, indicated by quite high ratings regarding instructional clarity ($M = 7.41$, $SD = 1.72$).

4. Discussion

This study empirically manipulated and inspected load-inducing factors from the long time postulated three-factorial framework (Sweller et al., 2011) simultaneously within a single experimental approach. In doing so, a working memory updating task (Ecker et al., 2010) was used. Due to the demand of remembering and recalling, such could be regarded as basal kind of learning task.

Overall, with increasing complexity extended reaction times and more errors occurred. Enhanced reaction times arose during both updating and recall under the presence of split attention. The effect of schema presence appeared in the case of errors in updating as well as recall phases, resulting in more errors without the opportunity to rely on previously exposed schemata. In the case of time, faster reactions with schema presence occurred only within updating steps. Contrary to the theoretically postulated independence of the outlined CL facets, some interactions between either complexity and split attention, complexity and schema presence or all factors could be observed in updating steps and final recall.

In terms of split attention, participants indeed had to cope with additional attentional demands at the outset of a trial. However, during updating steps, their focus persisted on just one spatial object at once, possibly indicating the absence of significant differences in errors in this phase. Additionally, the increase in reaction time in both update and recall steps with split attentional focus might have further compensated for errors. The lack of influence of split attention under low complexity for updating errors corresponds well with the outlined element interactivity effect (Chen et al., 2016; Sweller & Chandler, 1994). Such would state that learners' mental resources might be applicable for compensatory purposes in this case. Approaching the results on a neural level, in their research on mental rotation, Shepard and Metzler (1971) showed that an increase in the angle of rotation linearly aligns to an increase in reaction time. In a similar way, within the current experiment, an enhanced spatial distance between letters could have also resulted in an enhanced mental distance, potentially explaining the significant increase in reaction times in updating steps as well as final recall. On the other hand, spatial distance might have been helpful for some participants to mentally separate the letters. Such available mental space could have been used for constructing supportive letters in between to cope with the demanded transformations, possibly explaining the lack of significantly increased errors.

Converging the aspect of schemata, although configurations of letter sets were fully or partially repeated in schema-related trials, those could have been masked by the induced variation in updating transformations. The latter might have

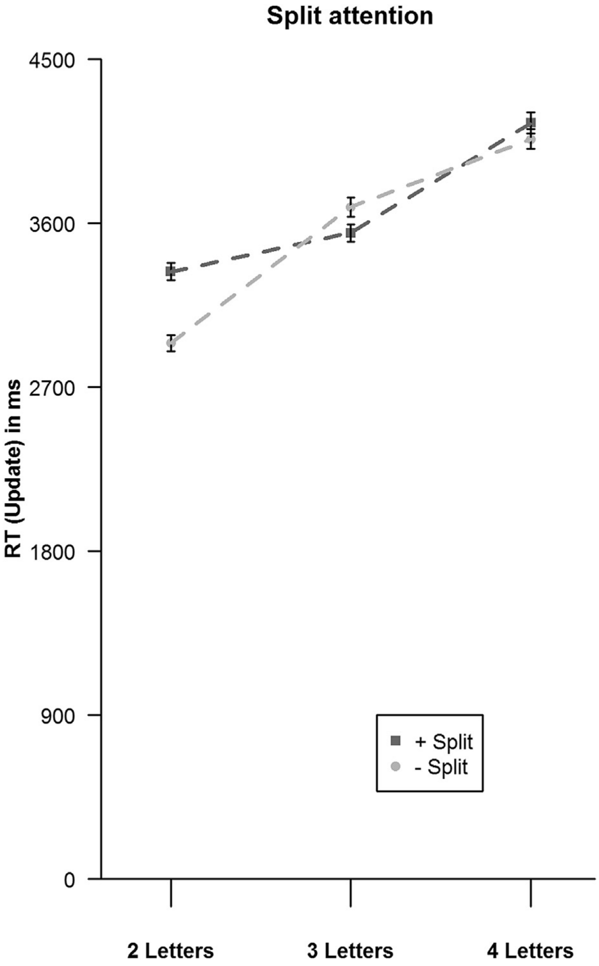

Fig. 3. Interaction of complexity and schema presence for RT_{recall}. Dots indicate mean values, error bars indicate standard errors and dashed lines were inserted to illustrate interactions.

increased cognitive demands since participants had to cope with interference resulting from former presentations of similar letter sets. Such mental operations could have put additional requirements on the anyway limited memory resources resulting in worse performance. Taking a separate look at different levels of complexity, especially under high complexity schemata held a compensatory influence when participants had to spread their attentional resources. By contrast, with focused attentional resources the benefit of inserting schemata became apparent just in the case of medium complexity.

Comparing effects for updating and recall steps, different patterns might have resulted due to the already outlined distinct sets of underlying mental operations. Referring to Ecker et al. (2010), updating comprises a set of features, each demanding a certain amount of time to be performed correctly. For this reason, with increasing effort via additional letters, larger spatial distance or the lack of schemata, participants needed more time to complete an updating step, apparent due to various significant effects in reaction time. By contrast, for final recall those transformations appeared just in an oblique manner, since correct recall requires correct updating beforehand, potentially explaining the overall lower reaction times during the recall phase.

Effects regarding individual aptitude and control variables outline the influence of concentrated attention on task performance. Moreover, besides of holding influence on the overall task performance, at least in some cases it seemed to affect how participants coped with increased mental demands due to raised complexity. This aligns to Schweppe and Rummer (2014), discussing the role of attentional focus in terms of cognitive resources. They postulate that participants with higher capacity exhibit more abilities to control attention and keep it focused on certain content. Taking a look at the development of task performance over time, obviously training effects occurred, resulting in faster and less erroneous responses the further people proceed in task completion. Such finding strongly indicates the development of overall task-related schemata that improve performance on the cognitive as well as motor level due to their both declarative and procedural nature (Gagné & Dick, 1983).

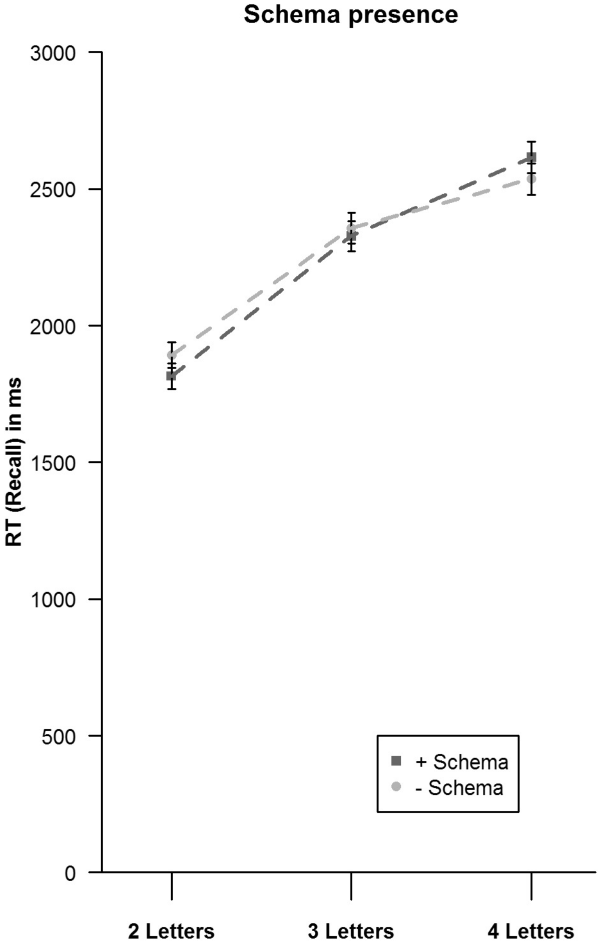

Fig. 4. Interaction of complexity and split attention for RT_{update}. Dots indicate mean values, error bars indicate standard errors and dashed lines were inserted to illustrate interactions.

4.1. Implications

Although independent effects of the CLT facets were postulated in advance due to the assumption of additivity (Sweller et al., 2011), the incidence of significant interactions points towards substantial overlap between those facets. On the one hand, such results are in line with Brünken, Plass and Moreno (2010) and Park (2010), indicating interference instead of pure additivity, and correspond well with recent reformulations of the theory (Kalyuga, 2011; Sweller, 2010). On the other hand, interactions on a statistical level might be distinct from substantial interrelations between facets on a task-related level. These may result from cognitive overload that explicitly arises from an unfavorable interplay of different load-inducing factors. In addition, difficulties in empirical separation might accrue since CL types reside on distinct levels of observation: Whereas ICL and ECL comprise structural characteristics related to content and presentation of a learning task, GCL involves processual features related to learning and knowledge acquisition. Such distinction aligns to diverse temporal perspectives within a learning task – momentary and short-term focused for ICL and ECL (learning input), but global and long-term focused for GCL (learning result), since building up schemata entails strong relations into long-term memory where knowledge can be stored permanently. Significant effects regarding concentration support the influence of individual aptitude variables, already indicated by Beckmann (2010) and Park (2010).

4.2. Limitations

Above all, the high level of task complexity could have weakened effects of split attention and schemata contributing to interindividual and intraindividual variance, obvious by small differences between conditions in the latter cases. Such assumption is supported by the high ratings regarding mental effort and task complexity within the concluding questionnaire. Moreover, complexity might have resulted not only by increasing the number of letters, but the letters itself for reasons of differences in familiarity throughout the alphabet and interindividual variations in associative connections. Such aspects

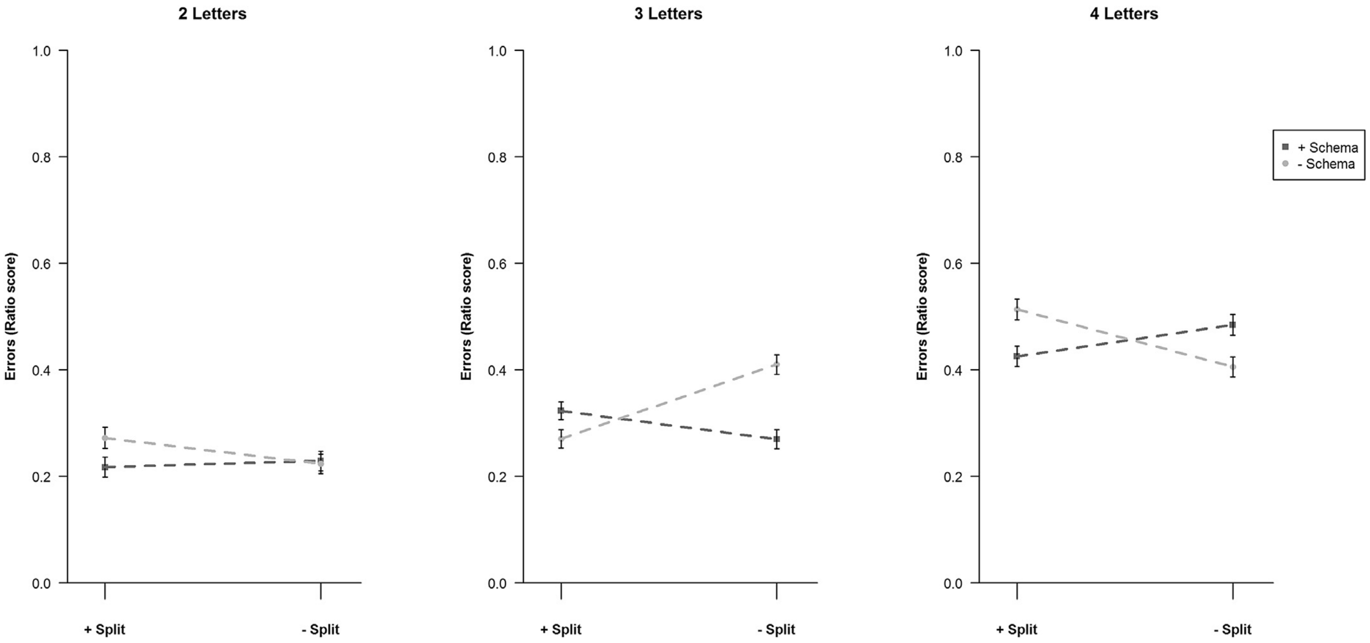

Fig. 5. Interaction of complexity, split attention and schema presence for Errors$_{update}$. Dots indicate mean values, error bars indicate standard errors and dashed lines were inserted to illustrate interactions.

Table 3
Standardized beta-coefficients, standard errors, t-values and significance levels of fixed effects in linear mixed model analyses for aptitude and control variables.

	$\text{RT}_{\text{recall}}$ (ms)				$\text{Errors}_{\text{recall}}$				$\text{RT}_{\text{update}}$ (ms)				$\text{Errors}_{\text{update}}$			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Main effects																
d2-R (*KL* score)[a]	−0.072	0.065	−1.109	.271	−0.206	0.049	−4.228	<.001	−0.147	0.057	−2.569	.012	−0.200	0.055	−3.617	.004
Task sequence[b]	−0.219	0.015	−14.268	<.001	−0.033	0.016	−2.062	.039	−0.114	0.016	−7.272	<.001	−0.027	0.016	−1.654	.098
Interaction effects																
Complexity × d2-R[b]	0.044	0.015	2.852	.004	0.012	0.016	0.718	.473	0.047	0.016	3.008	.003	0.004	0.016	0.259	.795
Split-attention × d2-R[b]	0.005	0.015	0.315	.753	−0.035	0.016	−2.165	.031	−0.011	0.016	−0.676	.499	−0.006	0.016	−0.387	.699
Schema presence × d2-R[b]	0.012	0.015	0.796	.426	−0.007	0.016	−0.406	.685	0.006	0.016	0.396	.692	−0.012	0.016	−0.721	.471

Note.
[a] *df* = 90.
[b] *df* = 2105. $\text{RT}_{\text{recall}}$ = reaction time during final recall, $\text{Errors}_{\text{recall}}$ = errors during final recall, $\text{RT}_{\text{update}}$ = reaction time during updating steps, $\text{Errors}_{\text{update}}$ = errors during updating steps. Values based on $N = 2208$ observations of $N = 92$ participants.

are prone to induce additional variance in task complexity that cannot be controlled in advance. Another potential confounding influence arises, for participants could have increased their sitting distance towards the screen to compensate for increased spatial distance between the letters. Furthermore, the increase in spatial distance depended on the amount of presented display objects inducing a huge gap particularly for two letters whereas distances in the case of four letters were considerably smaller. Due to these constraints, the demand to split up attention might have not been able to reach its full potential, bringing about minor effects as well as significant interrelations on both task and statistical level. Approaching the matter of schemata, for the outlined processual and long-term nature of schema acquisition, participants could have lacked resources to extensively engage in this process, resulting in small differences between conditions. In addition, the chosen manipulation might have directly contributed to increase previous knowledge, and thus has rather been an inherent part of the experienced task complexity. Such finding would further explain existing statistic interrelations between both facets.

4.3. *Prospect*

Due to its strong and masking effect, a predominant issue within following studies comprises the reduction of complexity. A distinction between low and medium levels of task difficulty might be more adequate to study instructional effects. In addition, more obvious opportunities to engage in schema acquisition should be included, for instance by applying support for coherence formation (Seufert & Brünken, 2006) during a longer practice sequence. Such would enable participants to build solid and elaborated relations within the presented instructional material. An alternative way of schema activation could involve variations in updating sequences. Regarding the aspect of split attention, the current study has raised the demand for validly inspecting effects of distance between elements to derive more systematic predictions on the amount of helping vs. harmful interspace within given learning material. Additionally, since the used learning material heavily relies on previous experience with the Latin alphabet, further studies might use alternative, culturally independent materials like abstract symbols. Even a different modality could be introduced, for instance via using simple sounds, either as stimuli or to indicate transformations.

5. Conclusions

Within media-related educational research, taking into account learners cognitive scopes and limitations constitutes a valuable approach with broad impact on the design of instructional material. Especially the theoretical concept of cognitive load described by the CLT exhibits a broad history of research in this field that has already provided insights for a variety of research questions. Nevertheless, in terms of the valid empirical assessment and interrelation of the theoretically described building blocks, there are still lots of open questions to be addressed in future research. The current study might be regarded as a small step contributing to this goal.

Acknowledgements

Research was conducted within the Research Training Group “CrossWorlds”, GRK 1780/1, funded by the German Research Foundation (DFG). Grateful thanks to Stefanie Dlouhy, Anna Dirk, Shirin Esmaeili Bijarsari and Daniel Wirzberger for their support during data collection and data preparation.

References

Beckmann, J. F. (2010). Taming a beast of burden – on some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction, 20*, 250–264.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2 – Revision. Aufmerksamkeits- und Konzentrationstest* [*Attention- and Concentration Test d2 - revised version*]. Göttingen, Germany: Hogrefe.

Brünken, R., Plass, J. L., & Moreno, R. (2010). Current issues and open questions in cognitive load research. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 253–272). New York, USA: Cambridge University Press.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). New York, USA: Cambridge University Press.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233–246.

Chen, O., Kalyuga, S., & Sweller, J. (2016). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, 1–13.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Erlbaum.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, MA, USA: Cambridge University Press.

Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*, 51–57.

Cowan, N., Morey, C. C., & Chen, Z. (2007). The legend of the magical number seven. In S. Della Sala (Ed.), *Tall tales about the mind and brain: Separating fact from fiction* (pp. 45–59). Oxford, U.K: Oxford University Press.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: an experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 170–189.

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Boca Raton, FL, USA: Taylor & Francis.

Gagné, R. M., & Dick, W. (1983). Instructional psychology. *Annual Review of Psychology, 34*, 261–295.

Garson, G. D. (2013). Fundamentals of hierarchical linear and multilevel modeling. In G. D. Garson (Ed.), *Hierarchical linear modeling: Guide and applications* (pp. 3–25). Thousand Oaks, CA, USA: SAGE Publications.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY, USA: Cambridge University Press.

Ginns, P. (2006). Integrating information: a meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511–525.

Irtel, H. (2007). *PXLab: the psychological experiments laboratory* (Version 2.1.19) [Computer software]. Mannheim, Germany: University of Mannheim. Retrieved April 30th, 2015 from http://www.pxlab.de.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*, 105–134.

Kalyuga, S. (2010). Schema acquisition and sources of cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 48–64). New York, NY, USA: Cambridge University Press.

Kalyuga, S. (2011). Cognitive load theory: how many types of load does it really need? *Educational Psychology Review, 23*, 1–19.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge, MA: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Nezlek, J. B., Schröder-Abé, M., & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung. Vorteile und Möglichkeiten der Mehrebenenmodellierung mit Zufallskoeffizienten [Multilevel analyses in psychological research. Benefits and chances of multilevel modeling with random coefficients]. *Psychologische Rundschau, 57*, 213–223.

Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology, 28*, 29–45.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71.

Park, B. (2010). *Testing the additivity hypothesis of cognitive load theory* (Doctoral dissertation). Retrieved September 14th, 2015 from http://scidok.sulb.uni-saarland.de/volltexte/2010/3478/pdf/Dissertation_BP_2010.pdf.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. Berlin, Germany: Springer.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (Version 3.2.0) [Computer software]. Retrieved April 23rd, 2015 from http://www.r-project.org.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508.

Schweppe, J., & Rummer, R. (2014). Attention, working memory, and long-term memory in multimedia learning: an integrated perspective based on process models of working memory. *Educational Psychology Review, 26*, 285–306.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*, 227–237.

Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology, 20*, 321–331.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Hove, East Sussex, U.K: Routledge.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY, USA: Springer Science + Business Media.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*, 270–285.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Upper Saddle River, NJ, USA: Pearson Education.