Acta Psychologica 179 (2017) 30–41

Contents lists available at ScienceDirect

Acta Psychologica

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

Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task

Maria Wirzberger*, Shirin Esmaeili Bijarsari, Günter Daniel Rey

Psychology of Learning with Digital Media, Institute for Media Research, Faculty of Humanities, Technische Universität Chemnitz, Germany

A R T I C L E I N F O

Keywords:
Cognitive load
Schema acquisition
Task complexity
Embedded interruptions
Performance monitoring

A B S T R A C T

Cognitive processes related to schema acquisition comprise an essential source of demands in learning situations. Since the related amount of cognitive load is supposed to change over time, plausible temporal models of load progression based on different theoretical backgrounds are inspected in this study. A total of 116 student participants completed a basal symbol sequence learning task, which provided insights into underlying cognitive dynamics. Two levels of task complexity were determined by the amount of elements within the symbol sequence. In addition, interruptions due to an embedded secondary task occurred at five predefined stages over the task. Within the resulting 2x5-factorial mixed between-within design, the continuous monitoring of efficiency in learning performance enabled assumptions on relevant resource investment. From the obtained results, a nonlinear change of learning efficiency over time seems most plausible in terms of cognitive load progression. Moreover, different effects of the induced interruptions show up in conditions of task complexity, which indicate the activation of distinct cognitive mechanisms related to structural aspects of the task. Findings are discussed in the light of evidence from research on memory and information processing.

1. Introduction

From a cognitive point of view, to inspect learning means to deal with schema acquisition as a relevant outcome. Since learning itself is a process and thus relates to the aspect of time, the need arises to inspect demands resulting from schema acquisition under a temporal perspective. Such has already been outlined by Renkl and Atkinson (2003) and extended in more recent research by Renkl (2014), in which distinct process stages are discussed. However, details on underlying progression models of schema acquisition have not yet been explicitly tested, although such knowledge would especially offer a benefit to multimedia-based learning scenarios. These settings are more prone to overload learners' mental facilities due to the multimodal, interactive and often temporally and spatially distributed presentation of information. Accepting the arising challenge, the research community needs to develop predictive models on opportune stages of task-related cognitive load to adapt instructional situations to learners' cognitive resource supply. The current study takes a step forward in clarifying extant theoretical assumptions on cognitive load by comparing plausible progression models on a statistical base.

A prominent cognitive theory, which provides advice for the conducive design of media-transmitted instructions, is the Cognitive Load Theory (CLT; Sweller, 1988; Sweller, Ayres, & Kalyuga, 2011). It is based on the assumptions of duration and capacity limitations in working memory, a virtually unlimited storage capacity of long-term memory and the representation and organization of knowledge via schemata. Learning performance, at a certain point in time, is impaired if the total amount of processing requirements exceeds the limitations of mental resources. According to previous research, cognitive load in learning situations arises from three different sources, which have to be considered on distinct observational and temporal levels. Firstly, task complexity in relation to learners' previous knowledge constitutes intrinsic cognitive load (ICL) as an inherent characteristic of relevant learning material (Sweller & Chandler, 1994). Secondly, the effects of inappropriate instructional presentation add to extraneous cognitive load (ECL), which is not related to relevant learning content. Both aspects affect performance on a more structural and short-term level. The aspect of ICL is traditionally defined in terms of element interactivity, characterized by the number of logically related information units (e.g., symbols, concepts, procedures), which learners have to process simultaneously in working memory (Sweller, 2010). ICL has been addressed experimentally by Beckmann (2010) and Wirzberger, Beege, Schneider, Nebel, and Rey (2016), who used a priori estimates of task complexity in arbitrary learning material. These estimates were based

* Corresponding author at: Psychology of Learning with Digital 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.actpsy.2017.07.001
Received 19 December 2016; Received in revised form 15 May 2017; Accepted 3 July 2017
Available online 12 July 2017
0001-6918/

on the number of interrelated dimensions or elements that participants had to deal with at the same time. By contrast, the conceptualization of ECL usually aligns with the violation of recommended multimedia design principles for presenting instructional content (Mayer, 2014; Sweller et al., 2011). Extending that view on the instructional situation as a whole, inappropriate situational constraints, which demand learners' mental resources, should also be taken into account (Wickens, Hollands, Banbury, & Parasuraman, 2013), for instance, when being interrupted during task execution. The arising task-irrelevant information represents a competing goal that detracts learners' cognitive resources from the actual task focus (Gerjets, Scheiter, & Schorr, 2003). In consequence, they might use less demanding but also less effective strategies to reach their learning goals. Thirdly, another source of cognitive load arises from the process of learning itself, specified as schema acquisition and automation within the theoretical framework (Kalyuga, 2010). Both aspects represent the germane cognitive load (GCL) and need to be considered in terms of processual and long-term accounts. This view corresponds to more recent approaches, which assume a dual framework of germane resources dealing with relevant aspects of instructional material and extraneous resources dedicated to handle irrelevant situational characteristics (Kalyuga, 2011; Sweller, 2010; Sweller et al., 2011). The authors postulate a sufficient approach to explain demands on learners' resources without redundancy, as GCL mainly reflects how learners deal with the amount of ICL imposed by a task. On the one hand, such reformulation respects the fact that certain cognitive load factors benefit learning, while on the other hand, it implies a highly motivated learner who is willing to spend all available cognitive resources on relevant aspects of the learning situation. Approaching GCL on a measurement level, changes in learning efficiency can be regarded as valid indicator of changes in the level of imposed load, since with increasing acquisition of knowledge structures the same performance can be achieved with less investment in cognitive resources (Sweller et al., 2011).

As already stated initially, cognitive schemata constitute an essential achievement of learning, since well-established and organized knowledge structures foster a fast and easy information retrieval. This raises the importance of inspecting underlying cognitive processes of schema acquisition in more detail. From a historical perspective, schemata can be described in terms of mental structures or networks of knowledge, stored in the long-term memory, which incorporate general representations of specific information about an individual's world (Bartlett, 1932). The core function consists of forming guidelines for the interpretation, categorization (Beck, 1964) and appropriate response towards any kind of sensory input (Bower, Black, & Turner, 1979; Neuschatz, Lampinen, Preston, Hawkins, & Toglia, 2002; Rumelhart, 1980). Gagné and Dick (1983) emphasize a more active view of schemata in terms of procedural rules related to the process of understanding. Anderson (1984) describes several functions of schemata, allocated to memory encoding on the one hand, and allocated to information retrieval on the other hand. Once established, schemata provide a considerable reduction in time and capacity needed for mental processing (Bransford & Johnson, 1972; Rumelhart & Ortony, 1977), since their use becomes increasingly automated (Shiffrin & Schneider, 1977). However, the use of schemata is prone to errors. In particular, inappropriate prior schematic knowledge can interfere with proper memory recall (Bartlett, 1932; Brewer & Treyens, 1981; Sulin & Dooling, 1974). Regarding structural issues, schemata comprise a set of non-identical units, which are interrelated in terms of shared similarities (Anderson, 1984; Bartlett, 1932; Rumelhart, 1980; Rumelhart & Ortony, 1977). They are usually characterized by chronological (Bartlett, 1932) and hierarchical (Rumelhart & Ortony, 1977) order, with sub-units relating to multiple larger schemata (Head & Holmes, 1911; Rumelhart & Ortony, 1977). Head and Holmes (1911) further postulated the adaptability and modifiability of schemata, meaning that smaller units can be interchanged or broken up. Piaget (1952) identified two mechanisms responsible for such alterations: assimilation incorporates new information into existing schemata when searching for relevant similarities, whereas accommodation expands existing schemata with new elements when detecting relevant differences. In a recent review, Ghosh and Gilboa (2014) summarized the broad historical literature on schemata and derived a set of necessary and additional features of cognitive schemata. Corresponding to the subsequently outlined overview, they emphasized associative network structures, the rest upon multiple episodes, a lack of unit detail and an adaptability to modifications as necessary features. Additional features comprise chronological relationships, hierarchical organization, cross-connectivity and embedded response options.

Referring back to the CLT perspective, as already outlined, constructing and storing schemata in long-term memory during the learning process imposes GCL (van Bruggen, Kirschner, & Jochems, 2002). Relevant cognitive load increases with effort invested in establishing and automating task-related schemata of knowledge (van Merriënboer, Schuurman, De Croock, & Paas, 2002). With increasing element interactivity in learning material and thereby imposed complexity, ICL increases and demands limited working memory capacity, as well as being responsible for keeping schema-relevant information present. As a consequence, with more interconnected elements represented in learning material, higher mental effort is necessary to maintain information and construct schemata. Arising demands can even prevent further construction of schemata, if complexity exceeds learners' available resources (van Bruggen et al., 2002). Already existing schemata can reduce complexity and thus cognitive load, by reducing the amount of information to be maintained in working memory. Moreover, elements stored in long-term memory can facilitate the effectively organized interpretation and storage of sensory input in relation to existing structures (Valcke, 2002). The importance of available schemata has further been shown by Pollock, Chandler, and Sweller (2002), who stated that mental load may impede any kind of learning, if prior knowledge from previously established basic schemata is lacking.

Besides these demands that inherently arise from the used learning material, unrelated situational characteristics can impact learning processes as well. For instance, being interrupted while performing a learning task represents a potential source of ECL, since it usually impairs learning performance and interferes with coherent schema acquisition (Mayer, 2014). According to Brixey et al. (2007), interruptions are defined as unplanned breaks in human activity, which are initiated by internal or external sources in a situated context and result in discontinuities in task performance. Such events are prone to reduce efficiency and productivity and contribute to errors. Related impairing factors as well as potential strategies of prevention have been broadly inspected by various researchers (e.g., Gillie & Broadbent, 1989; Monk, Trafton, & Boehm-Davis, 2008; Trafton, Altmann, Brock, & Mintz, 2003). A commonly used indicator to determine the disruptiveness of an interruption is the time needed to return to the suspended task. Trafton et al. (2003) refer to this period as *resumption lag*, which is usually characterized by an initial decrease in how quickly people can perform the interrupted task. Besides other factors, it is influenced by the duration of the preceding interruption, with increased interference by longer interruption durations (Monk et al., 2008). Referring back to instructional situations, apart from negative effects on learning, resumption performance can hint at the stage of schema acquisition at various points in time. Practically, learners' cognitive resources should be less affected by maintaining interrupted tasks when certain content has already been transferred from temporary working memory structures to more durable long-term memory structures. In this vein, interruptions induced at defined stages during a task can serve as a test of the “robustness” of acquired schemata over time.

Approaching temporal characteristics during schema acquisition in more detail, Leppink and van Merriënboer (2015) already suggested that it would be worthwhile monitoring performance and mental effort

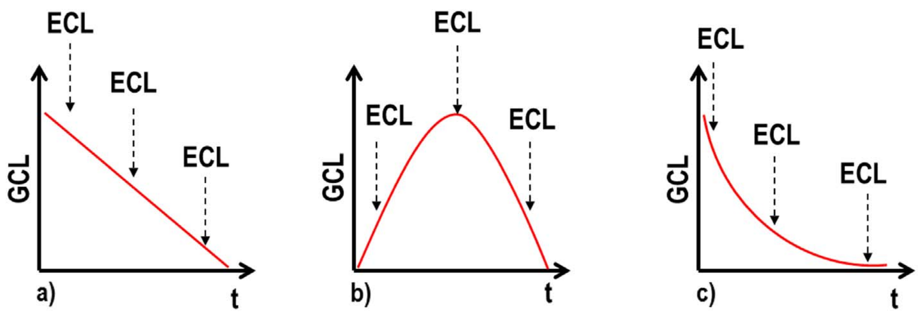

Fig. 1. Schematic outline of potential linear and nonlinear temporal models of cognitive load induced during schema acquisition. a) linear progression, b) quadratic progression, c) logarithmic progression.

continuously when performing repeated measurements within a learning task, as these aspects run through changes over time. Research on the expertise reversal effect and worked examples also outlined the need for load-reducing support in particular in the initial stages of a learning task (Kalyuga, 2007; Kalyuga, Chandler, Tuovinen, & Sweller, 2001). With increasing expertise - apparent from developed knowledge structures, which learners can rely on - load decreases and additional support becomes needless or even harmful (Rey & Buchwald, 2011). From this evidence, in the simplest case a decreasing linear trend in schema-related load progression could be assumed. However, as history of psychological research shows, trends related to cognitive processes are often not linear (Ebbinghaus, 1964; Yerkes & Dodson, 1908), raising the need of inspecting plausible nonlinear models of progression (see Fig. 1).

Following the work of Renkl and Atkinson (2003), most of the effort on building and developing schemata is likely spent during an intermediate stage, resulting in a peak in GCL embedded in an increasing and decreasing progression. This recalls the *quadratic* inverted U-shaped function on the relation between arousal and performance, as described by Yerkes and Dodson (1908). In a more recent article, Renkl (2014) specifies distinct phases in more detail. While the learners' goal is to become familiar with the basic declarative knowledge set related to the task domain in the first phase, they engage more actively in the establishment of knowledge structures in the second and third phases. During the last phase, due to the frequent application, the acquired schemata can be used automatically and more flexibly with minimal cognitive effort, which increases the robustness of the acquired cognitive skills.

An alternative progression, corresponding to this framework as well, is inspired by the well-known learning curve from Ebbinghaus (1964). Following an initially high investment of cognitive resources, fostering schema acquisition, with increasing establishment of schemata less load is put on the cognitive system, since pre-established schemata can be used. As already outlined, a lower level of load still persists due to automatization processes occurring with the frequency of schema use. This idea of a decreasing logarithmic progression of underlying cognitive resource demands receives support from worked-example research as well (Kalyuga et al., 2001): in the beginning, novices need to put a lot of effort in building stable knowledge structures, which lead to a changes from novice to expert status at some point in time in case of success. Expert performance is characterized by receiving maximal performance with minimal resource investment (Kalyuga, 2007), thus learners' cognitive resources should be demanded just to a minimal extent in that stage.

Summarizing the identified gaps in existing research, the study focused on changes in cognitive resource demands during a learning task, prompted by the acquisition and use of schemata (Bransford & Johnson, 1972; Rumelhart & Ortony, 1977). Since existing results on progressions in cognition-related processes indicate a nonlinear dynamic, a quadratic model (*hypothesis 1a*) and a logarithmic model (*hypothesis 1b*) are assumed to hold explanatory benefits over a strictly linear progression. Besides the process-related load component, the capacity devoted to deal with structural load features should change as well over time and be reflected in the way people cope with a given level of task complexity or interruptions embedded in the learning task. While effects of task complexity should affect performance at a general level, interruption effects should be reflected in both interruption and resumption phases (Foroughi, Werner, McKendrick, Cades, & Boehm-Davis, 2016; Monk et al., 2008). On this account, performance during resumption periods (*hypothesis 2a*) as well as performance during interruption periods (*hypothesis 2b*) are assumed to improve with increasing schema acquisition. Moreover, lower task complexity should result in better performance throughout all stages of the task (*hypothesis 2c*). A task setting with arbitrary learning material from basic cognition-related research should provide a concise and controlled opportunity to address underlying cognitive mechanisms, which might be further transferable to more complex and applied institutional and non-institutional learning scenarios.

2. Methods

2.1. Participants

A total of 116 undergraduate and graduate students from a midsized German university ($M_{\text{age}} = 23.25$ years, $SD_{\text{age}} = 4.34$, range: 18–44, 93 female) participated in the study. They were enrolled in Communication Sciences (59%), Psychology (24%), Education Sciences (8%) or other Social Sciences (9%), since the study was open to participants across the entire university. In terms of compensation for their participation, they received either a financial allowance of 5 € ($n = 36$) or course credits according to their curriculum ($n = 80$). When comparing sample characteristics between experimental conditions, neither group displayed significant differences in the distribution of age, $t(111.85) = 0.55$, $p = .581$, $d = 0.103$, gender, $\chi^2(1) = 0.00$, $p > .999$, disciplines of study, $\chi^2(3) = 2.39$, $p = .495$, or compensation, $\chi^2(1) = 0.12$, $p = .724$.

2.2. Design

The chosen learning task itself required participants to detect, remember and retrieve easy or difficult combinations of arbitrary geometric symbols, while being interrupted at several points in time by a visual search task. The acquired symbol combinations constituted the knowledge schemata that had to be obtained over the task. Within a 2 × 5-factorial mixed between-within design, task complexity was varied by the number of symbols that determine the following symbol (one vs. two). This factor represents the between-subjects ICL component and is addressed according to the outlined concept of element interactivity (Sweller & Chandler, 1994; Sweller et al., 2011). The interrupting visual search task characterizes the ECL component and was induced at five predefined stages during the learning task. This experimental manipulation aligns with the conceptualization of ECL, as indicated by Wickens et al. (2013). It was included as the within-subjects variable, while both structural load components were considered as independent variables in this setting. Learners' performance was recorded continuously across learning trials and interruptions via correctness and duration of responses to provide a constant assessment of changes in task-related demands. The resulting efficiency score reflects the mental resource investment pattern, which underlies the achieved performance (Sweller et al., 2011), and represents the GCL component as a dependent variable. As working memory capacity has been shown to moderate harmful effects of interruption (Foroughi et al., 2016), participants' overall mental resource capacity was derived as well. As such, shortened versions of two well-established working memory span tasks (Foster et al., 2015; Unsworth, Heitz, Schrock, & Engle, 2005; Unsworth, Redick, Heitz, Broadway, & Engle, 2009) were applied. A standardized questionnaire by Krell (2015) provided an additional examination of mental load and mental effort, whereas an open question on schema recall after the learning task enabled further insights into the quality of schemata acquisition over the task.

2.3. Materials

Computer-based tasks were realized with OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), operating on the Expyriment background (Krause & Lindemann, 2014), and provided on standard desktop computers with Windows 7 Professional 64 Bit, a 24″ monitor, a display resolution of 1920 × 1080 px and a video refresh rate of 60 Hz.

2.3.1. Schema acquisition task

The task on schema acquisition employed arbitrary learning materials to control for confounding effects of prior knowledge. The entire procedure comprised 64 trials, interrupted by a second task at five predefined points across the task. These interruptions occurred irregularly after a block of either eight or 16 trials (i.e., after trials 8, 24, 32, 40 and 56), to avoid predictability but appear at the same cognitive state in task routine for all participants. Responses and reaction times were recorded over the task for enabling continuous performance monitoring.

The main task required participants to detect, remember and retrieve interrelations between geometric symbols (circle, square, triangle and star). Interrelations were either simple (for example a circle resulted in a triangle, a square resulted in a star) or more complex (for example a square followed by a circle resulted in a square, but a circle followed by a square resulted in a star). As displayed in Fig. 2, at the outset of each trial, one or two symbols were presented for 2 s each, followed by a limited time span of 5 s to choose the subsequent symbol by mouse click out of four possibilities presented on the screen. These time spans align with the task setting used by Wirzberger et al. (2016), as well as a pretest with $N = 5$ participants ($M_{age} = 33.00$, $SD_{age} = 15.66$, range: 22–64, 3 female). The answer was followed by a feedback screen for 1 s, indicating the correctness of the response, with "Correct!" displayed in green for correct responses and "False!" displayed in red for false responses. In the case of a false response, the correct symbol was shown as well.

Interruption screens included a visual search task that was designed in line with results from interruption research, regarding the complexity of induced processing demands and the similarity to the main task (Gillie & Broadbent, 1989). Within a time span of 10 s, participants had to search, count and remember the amount of two indicated types of symbol from four different types of symbol being presented. As shown in Fig. 2, the chosen symbol stimuli comprised smaller versions of the simple geometric symbols used for the main task, similar to the stimuli set applied by Trick (2008), but without color. The task itself was inspired by evidence from the subitizing task (Jensen, Reese, & Reese, 1950) that suggested people using distinct mechanisms to discriminate smaller numbers (subitizing) and larger numbers (counting) of visually presented items. While original work claimed numbers up to and including six as subitizing range, more recent research has corrected this amount down to around three (Mandler & Shebo, 1982; Trick, 2008; Trick & Pylyshyn, 1994). However, this span might differ between individuals, depending on practice, age or cognitive skills in enumeration (Schleifer & Landerl, 2011; Trick & Pylyshyn, 1993) and can fully disappear in the presence of distractors (Trick & Pylyshyn, 1993). On this account, in the current task, seven to nine items for each target and distractor symbol were presented on the screen, to ensure that even participants from a student sample, used to perform complex cognitive operations, were required to invest substantial cognitive resources in the process of counting. Participants had to choose the correct numbers of the counted symbols, by mouse click one after another, from a set of potential numbers shown on the screen within 5 s per answer, thus the overall maximum interruption duration added up to 20 s. The chosen time spans were based on Wirzberger et al. (2016) and evidence from the pretest, as well as results introduced by Monk et al. (2008). They report a starting asymptote in resumption lag duration from interruption durations between 13 s and 23 s, indicating that further temporal increases in interrupting tasks would not crucially change the arising effect patterns.

2.3.2. Working memory span tasks

Two working memory span tasks were used prior to the schema acquisition task to obtain a baseline for participants' individual working memory capacity. They were based upon the shortened versions of the original Operation Span (OSPAN; Unsworth et al., 2005) and Symmetry Span (SSPAN; Unsworth et al., 2009) tasks from Foster et al. (2015). Both included a practice phase prior to the test trials, in which the participants had the opportunity to become familiar with each part of the task separately and in combination.

The OSPAN task consisted of five trials, including up to seven single letters from the Latin alphabet, which were randomly chosen out of a predefined set of 12 letters. They were shown one after another and had to be remembered by the participant. Each was preceded by a math problem, which was randomly chosen out of a pool of 192 potential math problems and had to be evaluated regarding its correctness. Each trial ended by choosing the correct order of the shown letters out of all 12 possible letters displayed on the screen by mouse click. Afterwards, participants received feedback on the percentage of correct evaluations for the math problems, as well as the correct recall of the letters. To ensure that they paid attention to both tasks equally, the task requested the percentage of correctly solved math problems to be kept at least at 85%, but to work as fast and precisely as possible at the same time.

Similar advice was given in the SSPAN task, which included four trials with up to five red squares in a 4 × 4 matrix, randomly chosen out of the 16 possibilities. Participants had to remember their position while dealing with a symmetrical picture before each matrix. Each picture was randomly chosen out of a pool of 48 pictures and required to evaluate its symmetry towards the vertical axis. Similar to the OSPAN task, after each trial, participants were required to indicate the positions in which the red squares had been shown in correct order by mouse click. Again, feedback was given concerning the percentage of correct answers for the symmetry pictures and the correct selections of the red square sequences.

2.3.3. Additional measures

After completing the task on schema acquisition, participants had to outline the assumed interrelations between the symbols on a separate sheet as free recall following an instruction. Moreover, to provide an additional measure of cognitive load, the task was followed by a paper-based questionnaire from Krell (2015) on experienced mental load and mental effort. While mental load refers to the amount of load arising from task and environmental demands, mental effort refers to cognitive capacity directly invested in dealing with the task (Paas & Van Merriënboer, 1994). The questionnaire comprised 12 items, six items for mental load (e.g., "The tasks were difficult to answer") and six items for mental effort (e.g., "I haven't taken particular trouble with the reply to the tasks"). Responses were given on a 7-point Likert scale, ranging from not at all (1) to moderately (4) and totally (7).

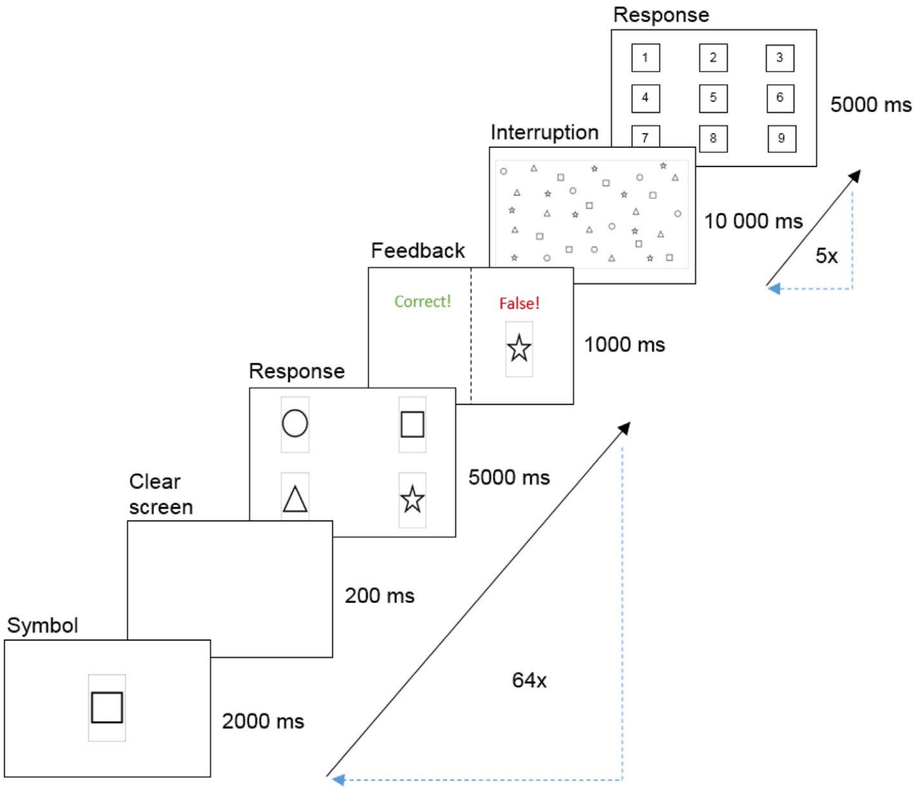

Fig. 2. Trial structure within the schema acquisition task. Presentation of second symbol in difficult condition analog with both symbols separated by an additional clear screen. Feedback screen according to participant's response (left half: correct choice, right half: false choice; translated version). Target symbols in interrupting task indicated by instruction above picture (not displayed). Response screen with instruction on target symbol to enter (not displayed), repeated for second target symbol with both screens separated by clear screen. Dashed lines inserted to emphasize repetition of trials.

2.4. Procedure

Sessions were conducted in a separate laboratory, equipped with 10 visually separated desktop computers for participants, which were arranged in rows of two, four and four. Up to six participants could participate per session with the experimenter always present at a separate desk in front. At the outset of each session, participants were welcomed and signed an informed consent, which outlined the purpose and procedure of the study and ensured that they were treated aligned with approved ethical standards and their privacy was respected. Afterwards, participants filled out a demographic questionnaire, while the experimenter started their first task. OSPAN and SSPAN appeared randomized as either the first or second task, whereas the subsequent task on schema acquisition was only randomized regarding its difficulty. Next, participants had to recall the obtained interrelations between the symbols on a separate sheet and fill out the questionnaire on cognitive load for the task on schema acquisition. In the end, participants were debriefed and approved.

2.5. Scoring

Efficiency scores from learning trials within the schema acquisition task were computed, following the likelihood model described by Hoffman and Schraw (2010), as a quotient of correct responses and reaction times within each trial. Since reaction times were retrieved in milliseconds, scores were multiplied by 1000 in order to obtain the proportion of correct responses per second. The resulting values indicate the use of available mental resources over the task and reflect the assumption that learners, which perform faster and less erroneous on the task, need to invest less mental capacities. Efficiency during interruptions was calculated in a similar manner, aside from summing up reaction times for the search and response parts of the symbol search task, resulting in smaller values due to longer overall time spans.

The partial load score for the OSPAN and SSPAN tasks was computed by awarding one point for each correctly recalled element in order to obtain a working memory span score from each task. This method of scoring was applied in line with Conway et al. (2005), who reported a clear advantage of partial credit scoring procedures over all-or-nothing scoring procedures.

For schema recall, sum scores were calculated on totally recalled sequences and correctly recalled sequences, resulting in values ranging from 0 to 4. For the questionnaire on mental load and mental effort (Krell, 2015), total scores were calculated by averaging items with regard to each factor. Three items per factor were reverse-coded and had to be recoded prior to aggregation.

3. Results

Three participants did not succeed in understanding the task and developing a task-related schema, given that they did not report anything in the final test on schema recall. As a consequence, they were excluded from subsequent analyses. When examining the influence of working memory span scores as a potential covariate, correlations between partial load scores from OSPAN and SSPAN tasks and the efficiency score did not indicate substantial relationships between both measures (efficiency score – OSPAN score: $r = .05$, $t(111) = 0.52$, $p > .05$; efficiency score – SSPAN score: $r = .07$, $t(111) = 0.74$, $p > .05$). In addition, no significant differences between conditions were indicated by the OSPAN score, $t(93.92) = -1.01$, $p > .05$, $1\text{-}\beta = .78$ (for $d = 0.50$ and $\alpha = .05$), nor by the SSPAN score, $t(105.95) = -0.19$, $p > .05$, $1\text{-}\beta = .84$ (for $d = 0.50$ and $\alpha = .05$). For these reasons, neither span score was included in the subsequent analyses.

To control for the interrupting potential of the used interruptions, two core findings from existing interruption research were explored prior to performing the main analyses. Firstly, there is prevalent evidence that interruption causes time costs, which is testable by comparing reaction times in trials directly before and after an interruption. Such was the case within the current task as well, given a significant main effect in an analysis of variance (ANOVA) of pre-post interruption comparison on reaction time, $F(1,112) = 34.59$, $p < .001$, $\eta_p^2 = .24$. Moreover, previous research found that the longer the duration of an interruption, the greater the time costs. This was also confirmed in this

Table 1
Comparison of tested conditional growth curve models with linear and/or non-linear predictors.

Model	Fixed effects	Random effects	*df*	BIC	cAIC	Δ_i	R^2
1	$Time_{lin}$ + Condition + $Interaction_{Time(lin)} \times Condition$	Slope $Time_{lin}$ + Participant Intercept	170.85	18450.47	18117.32	196.32	.303
2	$Time_{lin}$ + $Time_{quad}$ + Condition + $Interaction_{Time(lin)} \times Condition$	Slope $Time_{lin}$ + Participant Intercept	173.01	18316.29	17968.03	47.03	.317
3	$Time_{lin}$ + $Time_{log}$ + Condition + $Interaction_{Time(lin)} \times Condition$	Slope $Time_{lin}$ + Participant Intercept	173.37	18271.50	17921.00	0.00	.322

Note. Results based on $N = 113$ participants. lin = non-transformed linear variable, quad = quadratic transformed variable, log = logarithmic transformed variable, *df* = estimated degrees of freedom, BIC = Bayesian information criterion, cAIC = conditional Akaike's information criterion, Δ_i = cAIC difference, R^2 = conditional Pseudo-R^2 for GLMM.

setting, since interruption duration significantly predicted resumption duration in a linear regression analysis, $\beta = .11$, $t(563) = 6.35$, $p < .001$, and explained a significant proportion of variance in this variable, $R^2 = .07$, $F(1,563) = 40.35$, $p < .001$.

3.1. Inspection of load progression

Conditional growth curve models were computed in order to inspect load progressions over the task on a temporal perspective. They operated on the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2013) in R (R Core Team, 2016). Values for all relevant variables were z-standardized prior to their inclusion in the analyses to obtain standardized beta coefficients. Models were fit with restricted maximum-likelihood estimation and included time, condition and interaction between both predictors as fixed effects, and a time slope as well as subject-specific intercepts as correlated random effects. As outlined in Table 1, time was computed as either a linear, quadratic or logarithmic variable for fixed effects. For the purpose of comparing model performance, the Bayesian information criterion (BIC; Schwarz, 1978) and Akaike's information criterion (AIC; Akaike, 1974) comprise commonly used model selection criteria, with lower values indicating a better fit. Both take into account how well models fit to observed data, while simultaneously penalizing overly complex parameter structures (Kuha, 2004). Whereas the BIC focuses on identifying the "true" model, the AIC aims to predict new data and holds approximate equivalence to cross-validation procedures (Fang, 2011). Since especially the conditional AIC (cAIC) constitutes an adequate choice if a model includes meaningful random effects (Greven & Kneib, 2010), it was computed with the *cAIC4* package (Saefken, Ruegamer, Kneib, & Greven, 2014) in the present analysis. In addition, the conditional pseudo-R^2 for generalized linear mixed models, calculated with the *MuMIn* package (Bartoń, 2016) enabled the evaluation of model performance. Following conventions outlined by Burnham and Anderson (2002), when comparing cAIC differences between linear, quadratic and logarithmic models, the latter seemed superior, since only differences of $\Delta_i < 2$ indicate substantial empirical support. Fig. 3 shows the predicted changes in performance over the task and displays increased proximities between predicted and observed data points in terms of both quadratic and logarithmic progressions.

Standardized coefficients within the linear model yielded a medium-sized significant effect for the linear time predictor ($\beta = .31$ $SE = 0.02$, $t(111) = 19.24$, $p < .001$) and a smaller significant effect for the interaction between time and condition ($\beta = .05$, $SE = 0.02$, $t(111) = 3.03$, $p < .05$), whereas condition ($\beta = -.02$, $SE = 0.04$, $t(111) = -0.49$, $p > .05$) did not reach a significant effect.

Standardized coefficients within the quadratic model indicated a strong significant effect for the linear time predictor ($\beta = .78$, $SE = 0.04$, $t(3311) = 18.69$, $p < .001$) when including a quadratic time predictor, which achieved a medium-sized significant effect as well ($\beta = -.48$, $SE = 0.04$, $t(7005) = -12.22$, $p < .001$). In addition, the interaction between time and condition showed a small significant effect ($\beta = .05$, $SE = 0.02$, $t(111) = 3.03$, $p < .05$), whereas condition ($\beta = -.02$, $SE = 0.04$, $t(111) = -0.49$, $p > .05$) did not reach significance. Comparing linear and quadratic models with the χ^2 difference test indicated significant improvement due to the time quadratic predictor, $\chi^2(1) = 147.69$, $p < .001$, which increased the proportion of explained variance by 1.5%.

Standardized coefficients within the logarithmic model yielded significance for the logarithmic time predictor ($\beta = .32$, $SE = 0.02$, $t(7005) = 14.01$, $p < .001$) and the interaction between time and condition ($\beta = .05$, $SE = 0.02$, $t(111) = 3.03$, $p < .05$). The significant effect for the linear time predictor disappeared, when including the logarithmic time predictor ($\beta = -.02$, $SE = 0.03$, $t(743) = 0.80$, $p > .05$). As in the previous models, condition did not show a significant main effect ($\beta = -.02$, $SE = 0.04$, $t(111) = -0.49$, $p > .05$). Comparing linear and logarithmic models with the χ^2 difference test indicated significant improvement due to the logarithmic time predictor, $\chi^2(1) = 193.59$, $p < .001$, which increased the proportion of explained variance by 1.9%.

In summary, the findings reveal that participants in both conditions underwent changes in performance, but with several differences across distinct points in time. With reference to *hypotheses 1a* and *1b*, in terms of the progression model over time, both quadratic and logarithmic curves seem superior to the strictly linear progression. In particular the

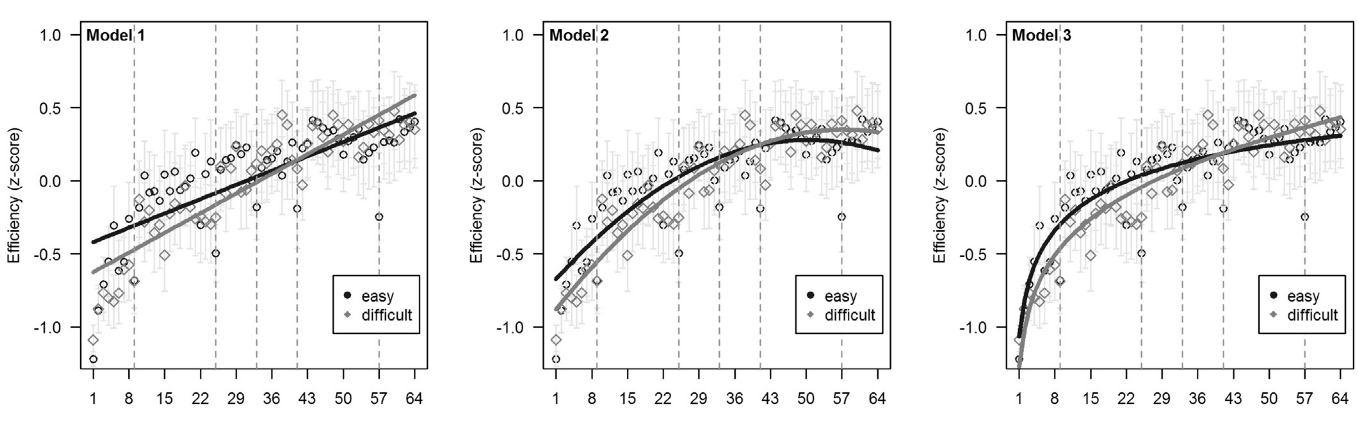

Fig. 3. Overall changes in efficiency across trials. Empty dots and rhombs show empirical mean values per trial, lines display predicted values for easy and difficult conditions from chosen models. Dashed vertical lines represent trials following an interruption. Error bars indicate 95% confidence intervals from empirical observations.

Table 2
Descriptive values of learning efficiency in pre-, post- and peri-interruption stages of the task.

Point in time	Pre-interruption performance		Resumption (post-interruption) performance		(Peri-)interruption performance	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
1	0.59 [0.44, 0.73]	0.40 [0.26, 0.54]	0.34 [0.22, 0.45]	0.33 [0.20, 0.46]	0.13 [0.09, 0.16]	0.10 [0.06, 0.13]
2	0.82 [0.65, 0.99]	0.56 [0.41, 0.72]	0.44 [0.33, 0.56]	0.59 [0.45, 0.73]	0.26 [0.21, 0.31]	0.25 [0.20, 0.30]
3	0.74 [0.60, 0.88]	0.78 [0.62, 0.94]	0.63 [0.50, 0.77]	0.81 [0.65, 0.96]	0.27 [0.23, 0.32]	0.30 [0.26, 0.35]
4	0.89 [0.73, 1.06]	0.82 [0.66, 0.98]	0.63 [0.47, 0.79]	0.79 [0.64, 0.94]	0.28 [0.23, 0.34]	0.34 [0.29, 0.38]
5	0.87 [0.71, 1.03]	0.93 [0.80, 1.07]	0.59 [0.45, 0.73]	0.98 [0.85, 1.11]	0.34 [0.28, 0.39]	0.36 [0.32, 0.41]

Note. Results based on $N = 113$ participants. Time in task coded according to point of occurrence. Cells display mean values [and 95% confidence intervals] in relevant trials before, after and during interruption for easy and difficult versions of the task.

logarithmic model holds substantial benefits, observable from model fits in BIC, cAIC, Δ_i and R^2, as well as the graphical impression.

3.2. Inspection of resumption performance

Reasonable differences in learning efficiency between conditions over the task, as well as due to the occurrence of interruptions (pre- vs. post-interruption comparisons), were already indicated on a descriptive level (Table 2). Values support a loss in performance due to being face with an interruption, especially for the easy task condition.

An ANOVA based on a linear mixed model was conducted with the *lmerTest* package (Kuznetsova et al., 2013) in R (R Core Team, 2016) to determine the influence of interruptions on learning efficiency over the task in both conditions. Due to findings from previous research and a significant negative correlation with efficiency, $t(1128) = -6.20$, $r = -18$, $p < .001$, interruption duration was included as additional fixed effect. Subject-specific random intercepts indicated the repeated-measures structure of the task, while pre- vs. post-interruption measurement and point of interruption occurrence over the task determined random slope components. Analyses supported the descriptive observations and revealed significant main effects regarding time of performance inspection, prior or after an interruption, $F(1,118.12) = 16.71$, $p < .001$, and over the task, $F(4,152.12) = 11.75$, $p < .001$. Moreover, significant interaction effects resulted regarding pre- vs. post-interruption depending on condition, $F(1,118.12) = 16.86$, $p < .001$, and time over task depending on condition, $F(4,152.12) = 11.75$, $p < .001$. Effects for condition, $F(1,111.02) = 0.62$, $p > .05$, interruption duration, $F(1,584.30) = 0.92$, $p > .05$, the two-way interaction between pre- vs. post-interruption and time over task, $F(4,554.71) = 0.85$, $p > .05$, as well as the three-way interaction between condition and both time-related factors, $F(4,554.71) = 0.85$, $p > .05$, failed to reach significance. The model achieved a conditional pseudo-R^2 of .442, revealing a substantial proportion of explained variance due to the included predictors. Statistically, the sufficient power of at least $1\text{-}\beta \geq .89$ for $\alpha = .05$ and $f = .25$ suggests the acceptance of the null hypothesis in all cases. However, to obtain additional support for the resulting nonsignificant results, Bayes factors were computed using the *BayesFactor* package (Morey & Rouder, 2015) in R (R Core Team, 2016). In brief, these values can provide a more in-depth inspection of competing hypotheses by specifying how much more times likely one is compared to the other (Dienes, 2014). By convention, a Bayes factor above the value of 3 can be taken as substantial evidence for the tested hypotheses, whereas values of $< 1/3$ should be considered as substantial evidence for the contrasting hypothesis (Jeffreys, 1961; Lee & Wagenmakers, 2014). When contrasting reduced models without the respective effect, representing the null hypothesis, with a full model including all tested effects, representing the alternative hypothesis, evidence resulted for omitting the effect of interruption duration, $BF_{01} = 7.090$ (error $\pm$ 1.29%), the two-way interaction between pre- vs. post-interruption and time over task, $BF_{01} = 1971.564$ (error $\pm$ 1.35%) and the three-way interaction between condition, pre- vs. post-interruption and time over task, $BF_{01} = 712.257$ (error $\pm$ 1.60%). By contrast, the obtained BF_{01} of 1.739 (error $\pm$ 1.71%) for the effect of condition did not clearly indicate an omission and suggested insensitive data.

Post-hoc pairwise comparisons using Tukey's HSD indicated significant pre-post interruption differences in efficiency for the easy condition ($p < .001$), whereas the difficult condition did not differ significantly ($p > .05$). In addition, regarding occurrences of interruptions over the task, significant differences in efficiency showed up in the easy condition between the first and third to last time points (each $p < .05$), and in the difficult condition between all five time points (each $p < .05$). Fig. 4 indeed indicates that changes in performance due to interruptions differ substantially between easy and difficult conditions at different points in time. Although participants suffered from interruptions throughout the entire task in the easy condition, performance seems to be rather unaffected by interruptions in the difficult condition. This impression is supported when taking a look at the amount of loss in efficiency, displayed in Fig. 5.

As such, *hypothesis 2a* is not supported, since efficiency clearly suffered from interruptions throughout the entire task in the easy condition, but seemed rather unaffected in the difficult condition. Moreover, the outlined condition-related pattern reverses the pattern postulated in *hypothesis 2c*.

3.3. Inspection of interruption performance

An ANOVA was conducted to inspect changes in interruption performance over the task, based on the same procedure as described for resumption performance, including subject-specific intercept as random effect. Interruption duration was inspected as an additional fixed effect as well, in line with existing empirical evidence and due to the significant negative correlation with efficiency, $t(563) = -9.16$, $r = -.36$, $p < .001$. Significant main effects for time of inspection, $F(4,464.77) = 12.53$, $p < .001$, and interruption duration, $F(1,546.95) = 12.55$, $p < .001$, were detected. The absence of significant results for the main effect of condition, $F(1,110.68) = 0.45$, $p > .05$, and the interaction between condition and time of inspection, $F(4,443.44) = 1.37$, $p > .05$, suggested an equal incidence of time-related patterns for both levels of difficulty. The model achieved a conditional pseudo-R^2 of .356, revealing a substantial proportion of explained variance by to the included predictors. Sufficient power followed from the given sample size ($1\text{-}\beta \geq .92$ for $\alpha = .05$ and $f = .25$), supporting the null hypothesis for nonsignificant effects. Again, to strengthen this assumption, Bayes factors were computed by comparing reduced models excluding this effects (null hypothesis) with the full model including all tested effects (alternative hypothesis). The resulting values favored the omission of the condition effect, $BF_{01} = 5.860$ (error $\pm$ 1.95%), but contradicted the omission of the interaction effect between condition and time of inspection, $BF_{01} = 0.085$ (error $\pm$ 0.63%). Fig. 6 supports this impression: although both conditions show comparable overall progressions, the observed levels in performance reverse after the second interruption.

Pairwise comparisons on the point in time, using Tukey's HSD, indicated highly significant differences between the first and all remaining interruptions (each $p < .001$), as well as between the second

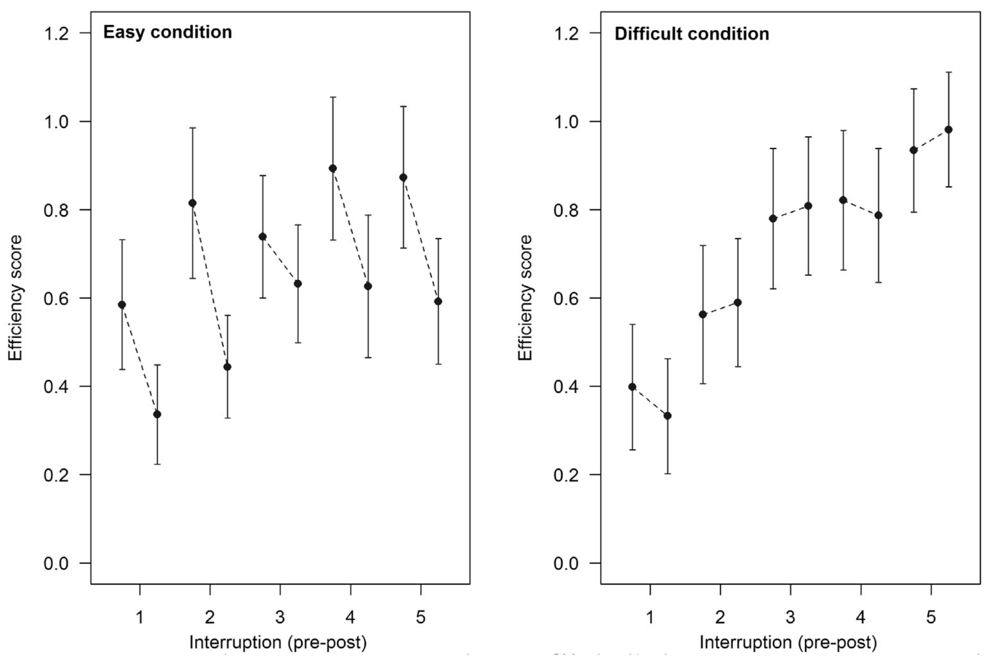

Fig. 4. Efficiency in trials immediately before and after an interruption. Dashed lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

and the fourth to last (each $p < .01$), and the third to last time points ($p < .01$). These results are already indicated on a descriptive level in Table 2.

From these observations, *hypothesis 2b* receives support, since performance increases over time in both conditions. However, since no significant difference was found between conditions, *hypothesis 2c* is not supported in terms of interruption performance.

3.4. Analyses of further variables

No significant overall differences between conditions of task difficulty showed up for experienced mental load, $t(110.62) = -1.52$, $p > .05$, and mental effort, $t(102.67) = -1.56$, $p > .05$. Statistically, for both analyses, the null hypotheses might be acceptable for an effect size of $d = 0.50$ because of sufficient power ($1\text{-}\beta = .84$, $\alpha = .05$). In addition, both groups achieved nearly equal scores on memorized schemata regarding the total amount of recalled relations, $t(79.93) = -0.78$, $p > .05$, as well as the proportion of them being correct, $t(111) = 0.88$, $p > .05$. The power within both analyses achieved .84 ($1\text{-}\beta$) for an effect size of $d = 0.50$ and $\alpha = .05$, suggesting the recommendation to accept the null hypothesis.

4. Discussion

The current study focused on the question concerning how load induced by schema-acquisition changes over time while concurrently taking structural load facets into account. Applying a basal symbol learning task, various levels of difficulty as well as interruptions at several points over the task were induced. Results indicate a nonlinear progression of schema-induced cognitive load over time that is influenced by the level of task complexity as well as the presence of interruptions. Harmful effects of interruptions were observed for the easy task condition, whereas none seemed to arise in the difficult task condition. In addition, interruption performance increased over time in

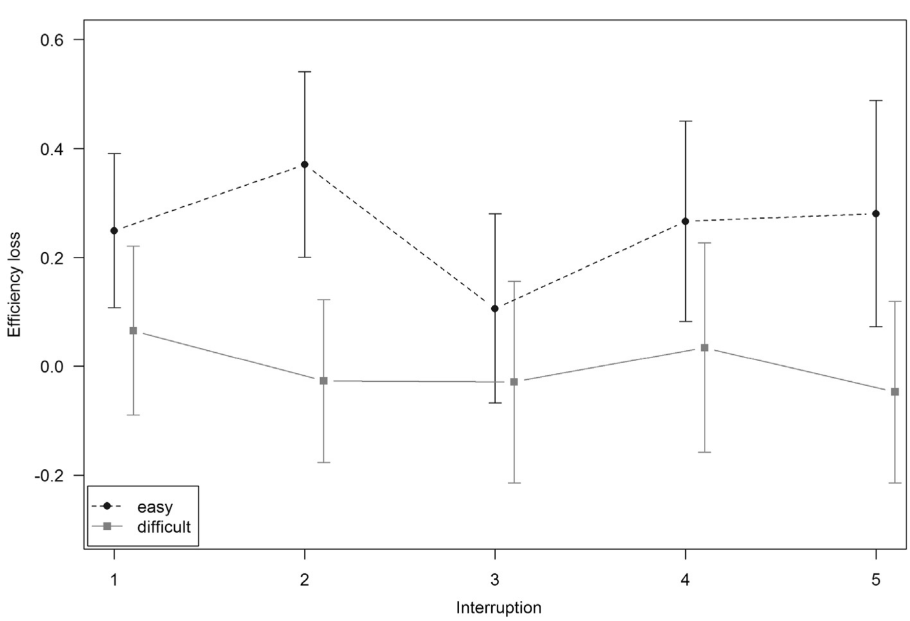

Fig. 5. Loss in efficiency due to interruption (pre-post comparison) in easy and difficult conditions. Lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

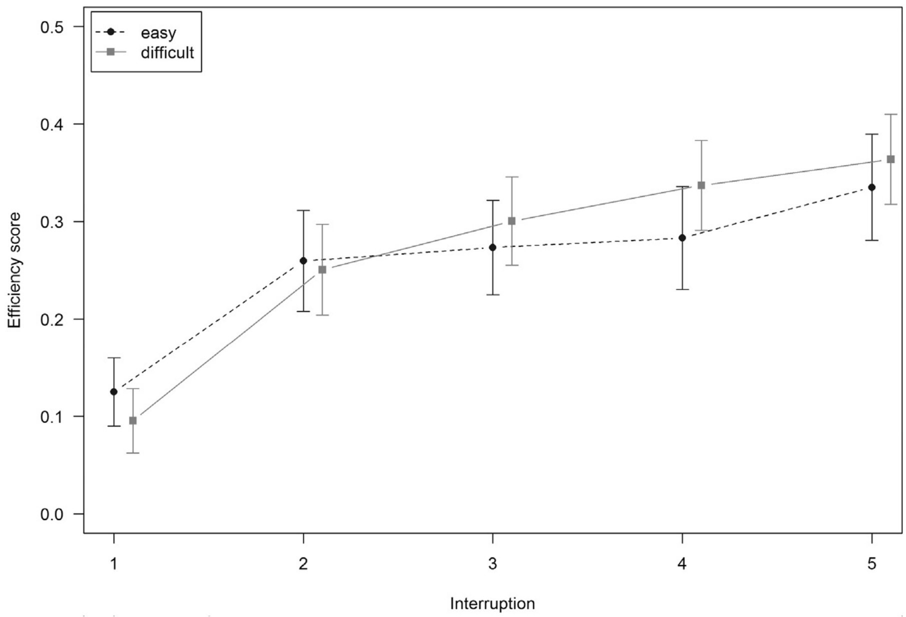

Fig. 6. Interruption performance over task. Lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

both conditions.

Regarding the obtained progression curves, the arising pattern of evidence supports recent theoretical reformulations of the initial framework of three cognitive load facets (Kalyuga, 2011; Sweller, 2010) and indicates that interactions between instruction- and content-related facets of learning situations and learners' cognitive resources go beyond purely additive relations. Moreover, the focus on temporal progression aligns with the initially introduced phases of skill acquisition and accompanying changes in learners' task-related focus (Renkl, 2014; Renkl & Atkinson, 2003). Taking a more detailed look at potential models of temporal progression, the explanatory benefit of nonlinear patterns of change aligns well with the described theories on learning and resource investment (Ebbinghaus, 1964; Yerkes & Dodson, 1908). In particular, the logarithmic model receives support from overall changes in performance over the task. However, for mapping performance on the initially postulated trend in cognitive load, scores have to be inverted and thereby provide only a rough estimation by indirect means. By contrast, in the case of the U-shaped model, the resulting tendency in performance directly mirrors the assumed progression in cognitive load. On the one hand, a pattern like this could indicate a higher selective investment of cognitive resources for establishing task-related schemata in the middle of the task. On the other hand, it might simply result from compensating for increased task demands during this period. As several ambiguities remain unsolved for both temporal models, the application of a continuous secondary task becomes necessary. Such task setting facilitates the examination of underlying cognitive resource distributions over time more directly on a measurement level. At a practical level, knowledge of progression patterns of learners' resource demand would provide hints for conducive guidance-fading procedures within the development of instructional materials (Sweller et al., 2011). Although there is significant evidence on the different effects of additional support, depending on learners' previous knowledge (expertise-reversal-effect; Kalyuga, 2007; Rey & Buchwald, 2011), more detailed insights into transition processes during learning tasks are still missing. A best-case scenario to address individual demands would comprise adaptive learning settings based on intelligent assistive technologies (Azevedo & Jacobson, 2008). However, if individual adaptations are not possible, due to the lack of technical resources, aligning the task design with a confirmed general model of cognitive load progression would provide a valid approximation. In the case of a logarithmic progression, a higher level of schema-related support should be included at the beginning of a task and fade towards the middle, whereas an inverted U-shaped progression would require a different approach. In the latter case, additional support should increase from the beginning, be available at the highest level around the middle of the task and decrease towards the end.

Approaching the observed differences between easy and difficult conditions in terms of resumption performance, the arising pattern agrees with the argumentation of Gillie and Broadbent (1989). The authors outline that when facing an interruption, to the same degree as low memory demands from the main task do not assure the *absence* of disruptive effects, high memory demands do not assure the *presence* of disruptive effects. Moving on to a more in-depth cognitive perspective, there might be distinctions in underlying memory processing, which correspond with differences between working memory and long-term memory systems. Whereas working memory is limited in both duration (about 20 s; Wickens et al., 2013) and capacity (about four elements; Cowan, 2010), long-term memory provides a virtually unlimited duration and capacity of information storage. On this account, working memory limitations are of minimal concern to learners whose knowledge in a domain is already well-established in long-term memory (Kalyuga, 2010). Applying this information to the current task setting, processing could have been limited to working memory resources within the easy task condition, since task demands fit to the available capacity. Thus, participants might have not felt the need to invest substantial cognitive resources into schema acquisition, as the elements only had a few possible combinations. This perspective goes beyond the pure resource-oriented view posed by the CLT, but takes into account learners' self-regulation abilities (Schwonke, 2015; Zimmerman, 1990), which actively control how available resources are invested during the learning process. By contrast, in the difficult task condition, people had to engage in memorizing and schema acquisition right from the outset, since task elements showed a higher variability of combinations. As a consequence, these participants might have put more effort into establishing knowledge structures and were able to access and adapt more easily to the changing content (Pollock et al., 2002; Valcke, 2002; van Bruggen et al., 2002). During resumption, previously developed structures could be retrieved, whereas participants relying on pure working memory resources had to rebuild all information from scratch. This might have resulted in a loss in performance. Another approach, addressing learners' self-regulation mechanisms, relates to volitional action control (Heise, Gerjets, & Westermann, 1997) and states that higher task difficulty prevents learners from being distracted by task-irrelevant information. This pattern arises from volitional protection of the main task goal against competing goal intentions from distracting information. Based on this theoretical framework, Scheiter, Gerjets, and Heise (2014) found impaired performance due to task-irrelevant information, but only for participants in easy task conditions, not in difficult task conditions. However, the authors showed that changes in performance were not significantly mediated by processing distracting information, which corresponds with equal performance in the interrupting task for easy and difficult conditions in the current study. Furthermore, following Csikszentmihalyi (1990), a sufficient level of complexity is required to foster participants' motivation to get involved in the task. This would provoke a higher resource investment, more focused attention and, in consequence, an increased level of task-related engagement that could enable faster automation of skilled performance. Although differences between both task conditions regarding learners' motivation seem plausible, such aspect was neither explicitly addressed in the task setting, nor available from experimenters' feedback. On this account, motivational explanations for the arising pattern of results are highly speculative, but including this aspect would provide a valuable extension within future studies. Referring to the experimental design, elements of symbol combinations in the difficult condition were presented separately, one after another, and were already "interrupted" by a clear screen. As such, participants in this condition might have been used to interruptions and also benefitted from extended presentation times. In consequence, they might have suffered less in performance. Moreover, the task only required to learn relatively few symbol combinations, allowing participants to apply more heuristic encoding strategies in the difficult condition after a while. Such provided the opportunity to increasingly speed up reaction times and thus achieve an overall enhanced efficiency in task performance. Additionally, with reference to Mandler and Shebo (1982), due to the relatively high number of target and distractor symbols, the interruption task could have triggered processes of estimation instead of mental counting, lacking the intended demands on cognitive resources as well. A potential disturbance related to the testing setting could have consisted in the form of time pressure via peer-induced stress, since participants had to keep waiting until each of them completed the last task. This situation may have forced slower participants to increase their speed and thus fostered a loss in concentration towards the end of the task setting, particularly under lower task demands. Additionally, the schema acquisition task was preceded by two tasks relying on working memory resources as well, giving way to possible mental fatigue or boredom at the end of the task.

Moreover, the chosen approach demonstrates the potential of interruptions for maintaining learners' active engagement in the task, obvious from the overall increase in interruption performance. This observation corroborates Trafton et al. (2003), who found that immediate interruptions without prior warnings become less disruptive over time. Comparable to effects of impaired text coherence (McNamara, 2001; McNamara, Kintsch, Songer, & Kintsch, 1996), the appearance of interruptions within different stages of the learning task seems to trigger a continuous state of active interference, fostering increased resource investment and resulting in deeper understanding. Further support results from research on desirable difficulties in learning by Bjork and Bjork (2011), discussing beneficial effects on retention and transfer when interleaved tasks require repeated reloading of memory content. For effectively studying such pattern, a sufficient length (Monk et al., 2008), complexity and similarity (Gillie & Broadbent, 1989) of the chosen interrupting task ensures that participants are required to break with the primary task. In consequence, since interruptions demand the use of long-term memory resources, their appearance should increase the durability of the learning content. However, the robustness of the acquired schemata was not explicitly addressed within the current task setting. It relates to questions about how long schemata are present in memory and to what extent they interfere with a new task. To this extent, a logical next step comprises to extend future studies by an evaluation of obtained schematic knowledge after a more extended test phase (Garner, Lynch, & Dux, 2016; van Merriënboer, Kester, & Paas, 2006). Potential transfer-related tasks might apply features like grouping or categorizing (Kalyuga, 2010). In addition, to gain better insights into task-adapted cognitive demands, future studies should monitor resource allocation in a more comprehensive way over the *entire* task, for instance, by applying a continuous dual-task setting or using psychophysiological measures. However, when adding a secondary task, it should be ensured that perception and response employ distinct modalities, compared to the primary task (Wickens, 2002), as valid predictions require resource interference to occur only at a cognitive level.

5. Conclusions

This work chose a concise and controlled approach from basic cognitive research to gain further insights into the temporal progression underlying schema-induced load. Results strongly indicate a nonlinear pattern of change over the task, which seems to be affected by structural, task-inherent characteristics as well. However, several issues related to underlying learner cognition remain unsolved within the current study and need to be addressed in more detail in future research. Despite some open questions, this framework comprises a promising way to approach existing "construction yards" and gain better insights into changing demands related to the process of schema acquisition in multimedia learning settings.

Acknowledgements

Research was conducted within the Research Training Group "CrossWorlds", funded by the German Research Foundation (DFG), GRK 1780/1. Grateful thanks to Anna Dirk and Stephanie Dlouhy for their help with data collection.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705.
Anderson, R. C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. Anderson, J. Osborn, & R. Tierney (Eds.), *Learning to read in American schools. Basal readers and content texts*(4th ed). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.
Azevedo, R., & Jacobson, M. J. (2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. *Educational Technology Research and Development*, *56*, 93–100. http://dx.doi.org/10.1007/s11423-007-9064-3.
Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, MA, USA: Cambridge University Press.
Bartoń, K. (2016). MuMIn: Multi-model inference. Retrieved from https://cran.r-project.org/web/packages/MuMIn/.
Beck, A. T. (1964). Thinking and depression. *Archives of General Psychiatry*, *10*, 561–571. http://dx.doi.org/10.1001/archpsyc.1964.01720240015003.
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. http://dx.doi.org/10.1016/j.learninstruc.2009.02.024.
Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY, USA: Worth Publishers.
Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220. http://dx.doi.org/10.1016/0010-0285(79)90009-4.
Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726. http://dx.doi.org/10.1016/S0022-5371(72)80006-9.
Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230. http://dx.doi.org/10.1016/0010-0285(81)90008-6.
Brixey, J. J., Robinson, D. J., Johnson, C. W., Johnson, T. R., Turley, J. P., & Zhang, J. (2007). A concept analysis of the phenomenon interruption. *Advances in Nursing Science*, *30*, E26–E42. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.217.3807.
van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representation of argumentation in CSCL and the management of cognitive load. *Learning and Instruction*, *12*, 121–138. http://dx.doi.org/10.1016/S0959-4752(01)00019-6.
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed). New York, NY, USA: Springer.
Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R.

W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. http://dx.doi.org/10.3758/BF03196772.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51–57. http://dx.doi.org/10.1177/0963721409359277.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY, USA: Harper & Row.
Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. http://dx.doi.org/10.3389/fpsyg.2014.00781.
Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Mineola, NY, USA: Dover Publications.
Fang, Y. (2011). Asymptotic equivalence between cross-validation and Akaike information criterion in mixed-effects models. *Journal of Data Science*, *9*, 15–21. http://dx.doi.org/10.6339/JDS.2011.09(1).652a.
Foroughi, C. K., Werner, N. E., McKendrick, R., Cades, D. M., & Boehm-Davis, D. A. (2016). Individual differences in working-memory capacity and task resumption following interruptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/xlm0000251.
Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226–236. http://dx.doi.org/10.3758/s13421-014-0461-7.
Gagné, R. M., & Dick, W. (1983). Instructional psychology. *Annual Review of Psychology*, *34*, 261–295. http://dx.doi.org/10.1146/annurev.ps.34.020183.001401.
Garner, K. G., Lynch, C. R., & Dux, P. E. (2016). Transfer of training benefits requires rules we cannot see (or hear). *Journal of Experimental Psychology: Human Perception and Performance*. http://dx.doi.org/10.1037/xhp0000215.
Gerjets, P., Scheiter, K., & Schorr, T. (2003). Modeling processes of volitional action control in multiple-task performance: How to explain effects of goal competition and task difficulty on processing strategies and performance within ACT-R. *Cognitive Science Quarterly*, *3*, 355–400.
Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema?: A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.010.
Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive?: A study of length, similarity, and complexity. *Psychological Research*, *50*, 243–250. http://dx.doi.org/10.1007/BF00309260.
Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, *97*, 773–789. http://dx.doi.org/10.1093/biomet/asq042.
Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, *34*, 102–254. http://dx.doi.org/10.1093/brain/34.2-3.102.
Heise, E., Gerjets, P., & Westermann, R. (1997). The influence of a waiting intention on action performance: Efficiency impairment and volitional protection in tasks of varying difficulty. *Acta Psychologica*, *97*, 167–182. http://dx.doi.org/10.1016/S0001-6918(97)00027-9.
Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist*, *45*, 1–14. http://dx.doi.org/10.1080/00461520903213618.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York, NY, USA: Oxford University Press.
Jensen, E. M., Reese, E. P., & Reese, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology*, *30*, 363–392. http://dx.doi.org/10.1080/00223980.1950.9916073.
Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539. http://dx.doi.org/10.1007/s10648-007-9054-3.
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. http://dx.doi.org/10.1007/s10648-010-9150-7.
Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*, 579–588.
Krause, F., & Lindemann, O. (2014). Expyriment: A python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*, 416–428. http://dx.doi.org/10.3758/s13428-013-0390-6.
Krell, M. (2015). Evaluating an instrument to measure mental load and mental effort using Item Response Theory. *Science Education Review Letters*, *2015*(1), 1–6 Retrieved from. (urn:nbn:de:kobv:11-100228733).
Kuha, J. (2004). AIC and BIC. Comparisons of assumptions and performance. *Sociological Methods & Research*, *33*, 188–229. http://dx.doi.org/10.1177/0049124103262065.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests in Linear Mixed Effects Models. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html.
Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, MA, USA: Cambridge University Press.
Leppink, J., & van Merriënboer, J. J. G. (2015). The beast of aggregating cognitive load measures in technology-based learning. *Educational Technology & Society*, *18*, 230–245. Retrieved from http://www.jstor.org/stable/jeductechsoci.18.4.230.
Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22. http://dx.doi.org/10.1037/0096-3445.111.1.1.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. http://dx.doi.org/10.3758/s13428-011-0168-7.
Mayer, R. E. (Ed.), (2014). *The Cambridge handbook of multimedia learning*(2nd ed). Cambridge, MA, USA: Cambridge University Press.
McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62. http://dx.doi.org/10.1037/h0087352.
McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43. http://dx.doi.org/10.1207/s1532690xci1401_1.
van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology*, *20*, 343–352. http://dx.doi.org/10.1002/acp.1250.
van Merriënboer, J. J. G., Schuurman, J. G., De Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*, 11–37. http://dx.doi.org/10.1016/S0959-4752(01)00020-2.
Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied*, *14*, 299–313. http://dx.doi.org/10.1037/a0014402.
Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. Retrieved from https://cran.r-project.org/web/packages/BayesFactor/index.html.
Neuschatz, J. S., Lampinen, J. M., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology*, *16*, 687–708. http://dx.doi.org/10.1002/acp.824.
Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371. http://dx.doi.org/10.1007/BF02213420.
Piaget, J. (1952). *The origins of intelligence in children*. New York, NY, USA: International Universities Press.
Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61–86. http://dx.doi.org/10.1016/S0959-4752(01)00016-0.
R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: Austria. Retrieved from http://www.r-project.org.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1–37. http://dx.doi.org/10.1111/cogs.12086.
Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, *38*, 15–22. http://dx.doi.org/10.1207/S15326985EP3801_3.
Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied*, *17*, 33–48. http://dx.doi.org/10.1037/a0022243.
Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension. Perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp. 33–58). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.
Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.
Saefken, B., Ruegamer, D., Kneib, T., & Greven, S. (2014). cAIC4: Conditional Akaike information criterion for lme4. Retrieved from https://cran.r-project.org/web/packages/cAIC4/index.html.
Scheiter, K., Gerjets, P., & Heise, E. (2014). Distraction during learning with hypermedia: Difficult tasks help to keep task goals on track. *Frontiers in Psychology*, *5*, 268. http://dx.doi.org/10.3389/fpsyg.2014.00268.
Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Developmental Science*, *14*, 280–291. http://dx.doi.org/10.1111/j.1467-7687.2010.00976.x.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. http://dx.doi.org/10.1214/aos/1176344136.
Schwonke, R. (2015). Metacognitive load – Useful, or extraneous concept? Metacognitive and self-regulatory demands in computer-based learning. *Educational Technology & Society*, *18*, 172–184. Retrieved from http://www.jstor.org/stable/jeductechsoci.18.4.172.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190. http://dx.doi.org/10.1037/0033-295X.84.2.127.
Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, *103*, 255–262. http://dx.doi.org/10.1037/h0036846.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. http://dx.doi.org/10.1207/s15516709cog1202_4.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. http://dx.doi.org/10.1007/s10648-010-9128-5.
Sweller, J., 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. http://dx.doi.org/10.1207/s1532690xci1203_1.
Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, *58*, 583–603. http://dx.doi.org/10.

1016/S1071-5819(03)00023-5.
Trick, L. M. (2008). More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception & Psychophysics*, *70*, 743–760. http://dx.doi.org/10.3758/PP.70.5.743.
Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 331–351. http://dx.doi.org/10.1037//0096-1523.19.2.331.
Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102. http://dx.doi.org/10.1037/0033-295X.101.1.80.
Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. http://dx.doi.org/10.3758/BF03192720.
Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654. http://dx.doi.org/10.1080/09658210902998047.
Valcke, M. (2002). Cognitive load: Updating the theory? *Learning and Instruction*, *12*, 147–154. http://dx.doi.org/10.1016/S0959-4752(01)00022-6.
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. http://dx.doi.org/10.1080/14639220210123806.
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.
Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016). One for all?! Simultaneous examination of load-inducing factors for advancing media-related instructional research. *Computers & Education*, *100*, 18–31. http://dx.doi.org/10.1016/j.compedu.2016.04.010.
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. http://dx.doi.org/10.1002/cne.920180503.
Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*, 3–17. http://dx.doi.org/10.1207/s15326985ep2501_2.