2019

Load-inducing factors in instructional design: Process-related advances in theory and assessment

Dissertation

submitted by

Maria Wirzberger

born on December 15, 1983
in Lohr am Main

for obtaining the academic degree
doctor rerum naturalium (Dr. rer. nat.)
in Psychology

at the
Faculty of Behavioural and Social Sciences
of the TU Chemnitz

Submitted: October 26, 2018
Defended: January 22, 2019

Advisors
Prof. Dr. Günter Daniel Rey
Prof. Dr. Josef F. Krems

The digital version of this document is available via

http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa2-327030

Without the knowledge of human cognitive processes, instructional design is blind.

(Sweller, Ayres, & Kalyuga, 2011, p. v)

Acknowledgements

Going through the process of this thesis has been a journey with many challenges, crises, and lots of emotional ups and downs. I am deeply grateful for the warm and kind support from all the people that helped me during this time. I could not have done this without you, you are my true heroes!

First of all, I thank my advisor Professor Günter Daniel Rey, who always trusted my ability to manage things, encouraged me to approach the next steps in my career ahead of time, was proud of my success, always had an open ear for my ideas, and never lost the kindness to deal with my worries. I also thank my co-advisor Professor Josef F. Krems, who provided just the right input at the right points, shared insights from his broad experience in cognitive research, and finally set the schedule to get this thesis finished. My very special grateful thanks go to Dr. Jelmer P. Borst, who made me feel very welcome during my research stay in Groningen, was a great supervisor, kindly supported me in completing the missing pieces for my thesis, introduced me to the joys of MATLAB, and prevented me from getting lost in the parameter setting desert of ACT-R. I further thank my “three musketeers” in Chemnitz, my valued colleagues and friends René Schmidt, Alexander Skulmowski, and Robert Herms, with whom I had the pleasure to do amazing research, participate in inspiring discussions, and who also pulled me out of the one or the other emotional crisis. Further, thanks to Shirin Esmaeili Bijarsari, who was a great and inspiring student assistant and co-author, supported me in conceptual issues, data collection, and manuscript revisions, and offered me her friendship. Thanks as well to my co-authors and colleagues Steve Nebel, Maik Beege, and Sascha Schneider, who provided critical comments on my studies and manuscripts that helped me to improve my work, and to Nico Münch, who always offered me kind and patient technical support with a smile on his face. Many thanks to my scholar Ronja Joschko, whom I had the pleasure to mentor for one year, for devoting her time to proofread my synopsis and her valuable advice for improvement. Moreover, I am very grateful for the opportunity to have been part of the Research Training Group “CrossWorlds” and thank the spokesmen Professor Maximilian Eibl and Professor Peter Ohler for accepting my application, appreciating my work, and providing me valuable career-related advice.

On my journey, I further had the pleasure to connect with great people who provided encouragement to pursue my goals. First of all, I give my special thanks to Dr. Falk Lieder, who challenged my mind, believed in my great potential, chose me to work with him, and thus motivated me to keep on track and cross the finish line. I further thank Professor Ute Schmid, who is a great mentor and role model, strengthened my trust in my abilities, and fostered the growth of my professional network. Grateful thanks to Professor Verena Nitsch, who guided my steps in organizing my first conference symposium, supported my career development from an early stage with her experience and advice, and encouraged me to reach for the stars. Moreover, special thanks to my academic advisors Professor Manfred Thüring and Professor Dietrich Manzey, who already provided encouragement during my master's degree, gave me helpful advice from the start of my dissertation project, and valued my scientific potential. I also thank Professor Nele Russwinkel, who kindly introduced me to ACT-R, established the contact to the Cognitive Modeling Group in Groningen, and fostered my personal growth during my master thesis. Many thanks to Professor Georg Jahn, who made me aware of the "hidden networks" in academia, appreciated my opinion and my ideas, and inspired me with his energy and enthusiasm. In addition, many thanks to Jun.-Professor Anne Kerstin Reimers, who supported me with her experience in the challenges of an academic career, her open ear for my worries, and her friendship.

I also want to thank a few people outside the "academic circle", who were real treasures on my way. First of all, very grateful thanks to Susanne Meisel, who gave me comfort with her friendship when I started to despair, always provided me an open ear and time for a walk and encouraged me to take a break from work sometimes. I further thank Elke and Frank Hiemann, who were great and caring neighbors that took interest in my progress and my life. I also thank Heinrich Bohaboj, who provided the necessary spiritual guidance to keep me connected to myself and my surroundings. Moreover, I thank Els van Dongen and Rudi Kamminga, who were lovely and encouraging hosts during my stay in Groningen.

Finally, I express my greatest, very special thanks to the most important people in my life. I am deeply grateful for all the emotional and organizational support that I received from my parents Bernadette and Stefan Wirzberger and my brother Daniel Wirzberger during this demanding and challenging time. Without all your great love and care, it would have been impossible to realize all these steps. Thank you for being part of my life!

Chemnitz, October 26, 2018

Maria Wirzberger

Summary

This thesis addresses ongoing controversies in cognitive load research related to the scope and interplay of resource-demanding factors in instructional situations on a temporal perspective. In a novel approach, it applies experimental task frameworks from basic cognitive research and combines different methods for assessing cognitive load and underlying cognitive processes. The first experimental study (N = 96) involves a basal learning task related to processes of working memory updating. Distinct facets of cognitive load are manipulated simultaneously with reference to the number, distance, and repetition of presented letters. Reaction times and errors in updating and recall steps of the task indicate the individual and combined influence of the varied features and individual aptitude variables and further emphasize the processual nature of schema acquisition. Within the second experimental study (N = 116) participants complete an abstract symbol learning task with different levels of task complexity according to the number of included elements. At five predefined stages over the task, interruptions are induced by an embedded visual search task. From the continuous monitoring of performance efficiency, a logarithmically decreasing change of invested cognitive resources seems plausible. Divergent effects of the induced interruptions related to conditions of task complexity hint on the activation of distinct cognitive strategies. By extending these behavioral results with a cognitive modeling approach based on the cognitive architecture ACT-R, underlying cognitive processes and mechanisms could be inspected in more detail. From the obtained insights, the potential for deconstructing and formalizing effects of increased task complexity on a cognitive level emerges. Furthermore, the time-related reconsideration of the cognitive load framework receives support on a neural level. The third experimental study (N = 123) involves a dual-task setting that requires participants to learn visually presented symbol combinations while memorizing auditory presented number sequences. Cognitive load during the learning task is addressed by secondary task performance, prosodic speech parameters, and physiological markers. In addition, the robustness of the acquired schemata is tested by a transfer task that requires participants to apply the obtained symbol combinations. The observed pattern of evidence supports the idea of a logarithmically decreasing progression of cognitive load with increasing schema acquisition. It further hints on robust and stable transfer performance under enhanced transfer demands. Taken together, the evidence obtained in this thesis emphasizes a process-related reconceptualization of the existing theoretical cognitive load framework and underlines the importance of a multimethod-approach to continuous cognitive load assessment. On a practical side, it informs the development of

adaptive algorithms and the learner-aligned design of instructional support and thus leverages a pathway towards intelligent educational assistants.

Zusammenfassung

Die vorliegende Dissertation nähert sich aktuellen Kontroversen in der Forschung zur kognitiven Beanspruchung in Lehr-Lernsituationen im Zusammenhang mit der Abgrenzung und dem Zusammenspiel ressourcenbeanspruchender Faktoren unter einer zeitbezogenen Perspektive. In einem neuartigen Forschungsansatz werden zu diesem Zweck experimentelle Aufgaben aus der kognitiven Grundlagenforschung angewendet und verschiedene Methoden zur Erfassung der kognitiven Beanspruchung und der Betrachtung zugrunde liegender kognitiver Prozesse kombiniert. Die erste experimentelle Studie ($N = 96$) beinhaltet eine basale Lernaufgabe im Zusammenhang mit Prozessen des Working Memory Updating. Definierte Facetten der kognitiven Beanspruchung werden simultan anhand der Anzahl, dem Abstand und der Wiederholung präsentierter Buchstaben manipuliert. Reaktionszeiten und Fehler in Updateschritten und finaler Wiedergabe im Zuge der Aufgabe zeigen den individuellen und kombinierten Einfluss der variierten Merkmale und individueller Charakteristika der Lernenden und unterstreichen zusätzlich den prozessualen Charakter des Schemaerwerbs. In der zweiten experimentellen Studie ($N = 116$) absolvieren die Teilnehmenden eine abstrakte Symbollernaufgabe mit unterschiedlichen Komplexitätsstufen, die durch die Anzahl der enthaltenen Elemente determiniert werden. Zu fünf vordefinierten Zeitpunkten im Aufgabenverlauf erfolgen Unterbrechungen durch eine eingebettete visuelle Suchaufgabe. Auf Basis der kontinuierlichen Erfassung der Leistungseffizienz erscheint eine logarithmisch abnehmende Veränderung der investierten kognitiven Ressourcen plausibel. Unterschiedliche Effekte der induzierten Unterbrechungen in den Bedingungen der Aufgabenkomplexität deuten auf die Aktivierung unterschiedlicher kognitiver Strategien hin. Mit der Erweiterung der verhaltensbezogenen Befunde um einen kognitiven Modellierungsansatz, basierend auf der kognitiven Architektur ACT-R, können die zugrunde liegenden kognitiven Prozesse und Mechanismen genauer untersucht werden. Die gewonnenen Erkenntnisse bieten das Potenzial zur Dekonstruktion und Formalisierung von Effekten erhöhter Aufgabenkomplexität auf kognitiver Ebene. Gleichzeitig stützen diese eine zeitbezogene Neubetrachtung des Rahmenmodells kognitiver Beanspruchung auf neuronaler Ebene. Die dritte experimentelle Studie ($N = 123$) nutzt einen Dual-Task-Ansatz, bei dem die Teilnehmenden visuell präsentierte Symbolkombinationen lernen, während sie sich gleichzeitig auditiv präsentierte Zahlenreihen merken sollen. Die kognitive Beanspruchung während der Lernaufgabe wird durch die Sekundäraufgabenleistung, prosodische Sprachparameter und physiologische Marker erfasst. Darüber hinaus wird die Robustheit der erworbenen Schemata durch eine Transferaufgabe geprüft, welche die Anwendung der zuvor erlernten Symbolkombinationen erfordert. Das

resultierende Evidenzmuster stützt die Idee eines logarithmisch abnehmenden Verlaufs der kognitiven Beanspruchung mit zunehmendem Schemaerwerb und deutet auf eine robuste und stabile Transferleistung auch unter erhöhten Aufgabenanforderungen hin. Zusammenfassend betonen die in der vorliegenden Dissertation gewonnenen Erkenntnisse eine prozessgeleitete Rekonzeptualisierung des bestehenden theoretischen Rahmenmodells der kognitiven Beanspruchung und unterstreichen zusätzlich die Bedeutung eines multimethodischen Ansatzes zur kontinuierlichen Erfassung der kognitiven Beanspruchung. Auf praktischer Seite lassen sich zentrale Hinweise für die Entwicklung adaptiver Algorithmen sowie eine an den Lernenden orientierte Gestaltung instruktionaler Prozesse ableiten, welche den Weg zu intelligenten Lehr-Lernsystemen eröffnen.

Overview on included original content

The following table lists the original content that is used in the indicated chapters of the submitted thesis.

Chapter	Original content
Article 1	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. doi: 10.1016/j.compedu.2016.04.010
Article 2	Wirzberger, M., Esmaeili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task, *Acta Psychologica, 179*, 30-41. doi: 10.1016/j.actpsy.2017.07.001
Article 3	Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications,* 3:46. doi: 10.1186/s41235-018-0138-z

Content

Acknowledgements i

Summary iii

Zusammenfassung v

Synopsis 1

1 Introduction 1

1.1 Practical significance 1

1.2 Theoretical background 2

1.3 Cognitive load assessment 4

1.4 Research focus 9

2 Experimental studies 9

2.1 Methods 10

2.2 Results 13

2.3 Implications 15

2.4 Limitations 16

2.5 Future work 17

3 Cognitive modeling 18

3.1 Model concept 21

3.2 Model comparison 24

3.3 Implications 29

4 Conclusions 30

References 32

Article 1 43

Abstract 43

1 Introduction 44

1.1 Cognitive Load Theory 44

1.2 Task complexity and ICL 45

1.3 Split-attention effect and ECL 46

1.4 Schemata and GCL 47

1.5 Working memory 47

1.6 The present study ... 48

1.7 Hypotheses ... 48

2 Methods ... 49

2.1 Participants ... 49

2.2 Design ... 49

2.3 Material ... 51

2.4 Procedure ... 53

2.5 Scoring ... 53

3 Results ... 54

3.1 Main effects ... 55

3.2 Interaction effects ... 56

3.3 Effects of aptitude and control variables ... 59

4 Discussion ... 61

4.1 Implications ... 63

4.2 Limitations ... 64

4.3 Prospect ... 64

5 Conclusions ... 65

Acknowledgements ... 65

References ... 66

Article 2 ... 71

Abstract ... 71

1 Introduction ... 72

2 Methods ... 77

2.1 Participants ... 77

2.2 Design ... 78

2.3 Materials ... 78

2.3.1 Schema acquisition task ... 78

2.3.2 Working memory span tasks ... 80

2.3.3 Additional measures ... 81

2.4 Procedure ... 81

2.5 Scoring 82
3 Results 82
3.1 Inspection of load progression 83
3.2 Inspection of resumption performance 86
3.3 Inspection of interruption performance 89
3.4 Analyses of further variables 90
4 Discussion 90
5 Conclusions 94
Acknowledgements 94
References 95
Article 3 103
Abstract 103
Significance 104
1 Background 105
1.1 Theoretical perspectives in cognitive load research 105
1.2 Approaches to cognitive load assessment 107
1.3 Present experiment 109
1.4 Hypotheses 109
2 Methods 109
2.1 Pre-study 109
2.1.1 Pre-study methods 110
2.1.2 Pre-study results 110
2.2 Main study 112
2.2.1 Participants 112
2.2.2 Design 112
2.2.3 Materials 113
2.2.3.1 Learning task 113
2.2.3.2 Schema application task 114
2.2.3.3 Questionnaires on retention and cognitive load 115
2.2.4 Procedure 115

2.2.5 Scoring 116
3 Results 117
3.1 Cognitive load progression 117
3.1.1 Primary task efficiency 118
3.1.2 Secondary task efficiency 118
3.1.3 Speech-related parameters 119
3.1.4 Physiological parameters 120
3.2 Retention and transfer performance 121
3.2.1 Retention performance 121
3.2.2 Transfer performance 121
3.3 Subjective cognitive load ratings 122
4 Discussion 123
4.1 Implications 124
4.2 Limitations 124
4.3 Future research 125
5 Conclusions 126
Acknowledgements 126
Declarations 126
References 128
List of tables I
List of figures II
Selbstständigkeitserklärung V
Curriculum Vitae VII
List of publications IX

Synopsis

1 Introduction

1.1 Practical significance

Recent advances in computer-based technology offer the potential to explore innovative solutions in learning and training contexts. The arising scope relates to various benefits from both the educating and the educated perspective – given that emerging challenges are considered in a sufficient way. When designing intelligent educational systems, the most important goal persists in providing each learner the opportunity to achieve the best possible learning outcome with appropriate effort. A sophisticated approach to enhanced cognitive skill acquisition can be achieved by tailoring instructional support to individual learners' needs, which also increases motivation and encourages sustained performance. For instance, during learning activities, an adaptive system could align the amount and speed of the presented content or the degree and scope of instructional feedback. At the same time, learners' cognitive resources should not be overloaded due to the variety of occupied modalities and provided interactive features. In consequence, for providing adequate feedback, such systems need sufficient input related to both performance and cognitive resource supply. While performance can be inspected via tracking learners' task-related progress, the actual pattern of invested cognitive resources needs to be derived from an enhanced scope of learner-related information. These can result from including additional channels, such as behavior or psychophysiological signals, as well as supporting evidence by computational models on task-related cognitive processes. Arising challenges in the development of adaptive educational systems firstly involve issues of adequate assessment. They address the accurate learner state recognition that requires intelligent algorithms for correctly interpreting the acquired signals. Secondly, system behavior needs to be adjusted continuously to meet learners' needs as sophisticated as possible. Motivated by both issues, this thesis explores the pattern of cognitive resource investment related to task performance by monitoring variations in cognitive demands over the task with a novel combination of sensitive measures related to performance, speech, physiological reactions, and computational cognitive modeling. On this account, it contributes evidence relevant to developing dynamic recognition algorithms underneath intelligent educational technologies.

1.2 Theoretical background

Approaching the subject on a theoretical level, instruction-related cognitive demands need to be considered and monitored carefully. A well-established theory in this field is the Cognitive Load Theory (Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merrienboër, & Paas, 1998), which addresses the construct of cognitive load in terms of working memory resources required to perform a certain task in a given situational context (Kalyuga & Plass, 2018). These demands relate to the ergonomic concept of strain (Beckmann, 2010; Kalyuga, 2011; Manzey, 1998), as they constitute a subjective experience that each individual learner has to cope with. The theory looks back on a history of about 30 years of active research with broad impact on conducive instructional design in a variety of domains. Upon its core assumptions, it resides on vested models of memory (Anderson, 1983; Atkinson & Shiffrin, 1971; Baddeley, 1992) that indicate limited working memory resources in terms of both duration and capacity of stored information. Besides a temporal duration of 20 to 30 s (e.g., Wickens, Hollands, Banbury, & Parasuraman, 2013), according to more recent research, the number of simultaneously available elements would reside around four (Cowan, 2010). By contrast, long-term memory resources provide nearly infinite storage capacity and duration and thus can be used to establish permanent knowledge structures. According to Schweppe and Rummer (2014), both memory systems are strongly interconnected, as working memory resources represent the activated part of long-term memory that holds the attentional focus. In line with evidence from schema theory (Anderson, 1983), the emerging organized knowledge structures are described as schemata that involve both declarative and procedural components (Gagné & Dick, 1983). Existing schemata influence how learners manage certain learning content and can be modified with new knowledge. In a recent review summarizing the evidence on schemata, Gosh, and Gilboa (2014) describe associative network structures, the foundation in multiple episodes, a lack of unit detail, and adaptability as core characteristics of the schema concept. With reference to established models of learning and skill acquisition (Anderson, 1982; Fitts & Posner, 1967; Ebbinghaus, 1964) higher amounts of resource investment are plausible in earlier process stages when knowledge structures still have to be established. The initially declarative knowledge becomes increasingly procedural and automated with task progression and in consequence, demands less cognitive resources.

Since its first description in the 1980s, the Cognitive Load Theory underwent several stages of refinement in terms of the postulated facets of cognitive load in learning situations. In the beginning, it focused on the prevention of harmful effects from task-irrelevant aspects, referred to as extraneous cognitive load (Sweller, 1988). This kind of resource demands arises from an inappropriate instructional presentation, for instance, due to inherent demands to split the attention between relevant sources of information (Ginns, 2006; Schroeder & Cenkci, 2018). On a broader level, it further involves interfering situational aspects of the learning context such as the prevalence of competing goals (Gerjets, Scheiter, & Schorr, 2003) by additional or interrupting tasks. The next stage of development expanded the focus on a load-inducing facet relevant to the learning task: the complexity of the used learning material in relation to existing previous knowledge (Sweller & Chandler, 1994). Along with defining this facet as intrinsic cognitive load, the concept of element interactivity was introduced, which emphasizes the interrelation of information elements as source of complexity (Chen, Kalyuga, & Sweller, 2017; Ngu, Phan, Yeung, & Chung, 2018). Characteristically, changes in element interactivity are related to the nature of what is learned (Sweller, 2010) and its amount should be kept at a manageable level for the individual learner to foster optimal learning outcomes. On this account, intrinsic cognitive load offers a toehold for adaptive learning procedures tailored to learners' expertise. A step ahead, in addition to extraneous and intrinsic cognitive load, a further source of cognitive load with beneficial effects for learning was introduced, primarily on theory-based accounts to explain so far unexplainable patterns of evidence (Paas, Tuivonen, Tabbers, & Van Gerven, 2003; Sweller et al., 1998). This so-called germane cognitive load emerges from learning-related processes of schema acquisition and automation and operates under the assumption of highly motivated learners that devote all available resources to these processes. Extraneous, intrinsic, and germane cognitive load were assumed to contribute independently and additively to overall cognitive resource demands in learning contexts (Sweller et al., 1998). This triarchic model of cognitive load can further be connected to the Cognitive Theory of Multimedia Learning (Mayer, 2009), another influential explanatory framework in instructional design. Although there is no exact mapping, essential cognitive processing relates to intrinsic cognitive load, as it deals with the selection and representation of relevant learning material in working memory. Extraneous cognitive processing is caused by suboptimal instructions, reminding of extraneous cognitive load, while generative cognitive processing corresponds to germane cognitive load by an active organization and integration of learning contents as well as learners' level of motivation (Kalyuga, 2011; Mayer, 2009).

Although the outlined facets have been broadly discussed in the corresponding literature, issues regarding the proper empirical assessment and psychometric separation persist. Such raise doubts on the originally postulated assumption of their purely additive interplay, as well as the independence of the later introduced germane cognitive load facet. To address the arising concerns, suggestions emerged to reformulate germane cognitive load as germane resources invested to deal with task-relevant intrinsic cognitive load (Kalyuga, 2011; Sweller, 2010, 2018) contrary to extraneous resources to deal with task-irrelevant extraneous cognitive load. Kalyuga and Singh (2016) argue even more towards a strict re-reduction of the framework into a two-component model that merely differentiates facilitative (productive) and impairing (unproductive) cognitive load factors and completely subsumes germane cognitive load under the facet of intrinsic cognitive load. By contrast, Seufert (2018) emphasizes the reasonability to retain the separation of intrinsic and germane cognitive load when considering aspects of self-regulation. In this context, both intrinsic and extraneous cognitive load facets represent task affordances imposed by the learning material, whereas germane cognitive load refers to learner-based decisions. She further criticizes the mainly static and deterministic perspective on cognitive load and clearly outlines the benefits of a more dynamic view on changes in cognitive load during learning. Following de Jong (2010), the separate consideration of the facets of intrinsic and germane cognitive load receives further confirmation as both represent distinct ontological categories. Whereas intrinsic cognitive load is related to the static complexity of the presented material, germane cognitive load refers to dynamic cognitive processes. Galy, Cariou, and Melan (2012) also outline the asymmetric nature of these facets that are supposed to act on different components of the cognitive system. According to Schnotz and Kürschner (2007), decreasing levels of cognitive load with increasing expertise are indeed plausible, which further advocates to adopt a process perspective on cognitive load, as claimed by Beckmann (2010).

Taken together, the existing literature reveals the lack of a time-related perspective in instructional cognitive load research. Such drives the demand of a processual reconceptualization of the three-component model that quantifies temporal changes resulting from schema-acquisition across the task. Based on this position, as documented in Wirzberger, Beege, Schneider, Nebel, and Rey (2016), Wirzberger, Esmaeili Bijarsari, and Rey (2017), and Wirzberger, Herms, Esmaeili Bijarsari, Eibl, and Rey (2018), the current thesis follows distinct levels of inspection of the outlined cognitive load facets with focus on their interplay during the learning process. Inspired by the concept of the zone of proximal development (Vygotski, 1963), the goal of intelligent adaptive systems would then be to keep the resulting cognitive load pattern at a manageable level for each learner at all stages.

1.3 Cognitive load assessment

According to Sweller (2018), the Cognitive Load Theory has originally been developed as a theoretical construct to explain experimentally obtained results, with little attempt to actually measure cognitive load. Nevertheless, since its initial description a variety of cognitive load measures has emerged (Paas et al., 2003; Sweller et al., 2011; Zheng, 2018). They operate on various parameters that can be categorized into subjective ratings, performance measures, physiological markers, and behavioral indices. Following Chen et al. (2016), while performance measures directly reflect task-related outcomes, behavioral indices hold information that does not directly affect domain-based outcomes. Contrary to physiological markers, the occurrence of behavioral indices can mostly or entirely be controlled by the learner. Related to arising differences in learners' evaluation of the complexity of the presented content and the benefit of the provided instructional support with increasing schema acquisition (Martin, 2018; Schnotz & Kürschner, 2007), a continuous monitoring of cognitive resource demands is advisable, which constitutes a core focus of the current thesis. An overview of the cognitive load indicators used in the related experimental studies is provided in *Table 1*.

Table 1

Cognitive load measures applied by the experimental studies in the included articles

Study	Subjective ratings	Performance	Physiology	Behavior
1	Paas (1992)	Reaction times Error rates	-	-
2	Krell (2015)	Efficiency Interruptions	-	-
3	Leppink et al. (2013)	Efficiency Secondary task	HR, SCR	Prosody

Note. Efficiency = correct responses per second, HR = heart rate, SCR = skin conductance response, Prosody = number and duration of silent pauses, phoneme-based articulation rate.

Amongst the earliest attempts to provide insights into cognitive demands arising from learning situations, subjective rating scales comprise a broadly used instrument. In particular, the unidimensional mental effort scale by Paas (1992) offers a convenient and easily usable option, although its informative value as single rating is limited. It requires participants to rate their perceived mental effort on a nine-point Likert scale ranging from "very, very low" to "very, very high". Kalyuga, Chandler, and Sweller (1999) use a modified scale to assess subjective mental load with a rating of instructional difficulty on a seven-point Likert scale ranging from "extremely easy" to "extremely difficult". In line with this procedure, Wirzberger et al. (2016) applied the mental effort scale accompanied by a rating on estimated task difficulty to enhance the scope of the stated predications. A more differentiated questionnaire that aims at assessing experienced mental load and mental effort is provided by Krell (2015). While mental load refers to cognitive demands arising from task-related and situational characteristics, mental effort refers to cognitive capacity invested in dealing with them. The questionnaire involves six items for each mental load (e.g., "The tasks were difficult to answer.") and mental effort (e.g., "I have made an effort at the processing of the tasks.") that have to be rated on a seven-point Likert scale from "not at all (1)" to "moderately (4)" and "totally (7)". It was used by Wirzberger, Esmaeili Bijarsari, et al. (2017) to obtain additional insights in invested cognitive resources across conditions. Contrary to the previously reported ratings, the questionnaire developed by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013) addresses the facets of intrinsic, extraneous, and germane cognitive load separately. It includes three questions on each intrinsic and extraneous cognitive load and four questions referring to germane cognitive load. The first two categories are closely connected to the underlying conceptual definitions by tapping either the complexity of the topic to be learned or the clarity of the instructional explanations (Ayres, 2018). In terms of the latter category, the related questions focus on understanding and knowledge acquisition, which turned out to be a rather controversial issue due to the lack of meaningful results in some studies (e.g., Leppink, Paas, Van Gog, Van der Vleuten, & Van Merriënboer, 2014). The reversed effect pattern for the germane cognitive load facet reported by Wirzberger et al. (2018), who also applied this questionnaire, further supports this critique. Leppink et al. (2014) discuss the related difficulties with reference to the already outlined re-reduced two-factorial cognitive load model. A more recent questionnaire by Klepsch, Schmitz, and Seufert (2017) addresses this issue by including the effort component more explicitly in questions related to germane cognitive load. The authors still emphasize the value of the three-factor model of cognitive load facets from a measurement point of view and state a more general applicability of their questionnaire across a wider range of educational subjects and domains.

Cognitive resource demands are further reflected in a more indirect way in performance-related parameters such as reaction times, error rates, and accuracy. These measures have a broad application and long history of use across a variety of psychological fields and disciplines. According to relevant literature from instructional cognitive load research (Hoffman & Schraw, 2010; Paas & Van Gogh, 2006) opposed to single accuracy or reaction time measures, a combined metric can be used as indicator for the quality of acquired cognitive schemata and thus offers a higher indicative value. Hoffmann and Schraw (2010) compare different approaches for calculating efficiency scores from both performance and effort indicators and outline the dependency of the chosen measure on the nature of the research question and the construct of interest. On this account, Wirzberger, Esmaeili Bijarsari, et al. (2017) and Wirzberger et al. (2018) applied an efficiency measure calculated from correct responses and reaction times according to the likelihood model, to obtain insights in changes across the process of schema acquisition. On methodological accounts, secondary tasks constitute a sophisticated way to shed light on task-related cognitive resource demands. They operate on the rationale that changes in working memory load related to a primary task can be monitored by a secondary task (Sweller, 2018; Kraiger, Ford, & Salas, 1993) and have already been proven reliability and validity in cognitive load assessment (e.g., Korbach, Brünken, & Park, 2017; Park & Brünken, 2018). This measurement approach is inspired by existing dual-task paradigms that apply a variety of tasks ranging from counting or reciting the alphabet to finger tapping, and humming a melody (e.g., D'Espositio, Onishi, Thompson, Robinson, Armstrong, & Grossman, 1996). Already applied secondary tasks in learning contexts include the observation of changes in auditory or visual stimuli (Brünken, Plass, & Leutner, 2004; Brünken, Steinbacher, Plass, & Leutner, 2002), requirements to memorize additional content (Renkl, Gruber, Weber, Lerche, & Schweizer, 2003; Wirzberger et al., 2018), the classification of auditory stimuli while performing a motor learning task (Esmaeili Bijarsari, Wirzberger, & Rey, 2017), or the performance of motor tasks such as tapping a previously learned rhythm by foot (Park & Brünken, 2015). However, choosing an appropriate secondary task that neither interferes with primary task requirements nor lacks sensitivity to observe arising demands still comprises a challenge when applying such task procedure continuously.

Increased cognitive demands also affect physiological states such heart rate, skin conductance, or brain blood flow dynamics. Due to the related characteristic of continuous assessment, they are particularly suited to obtain temporal progression patterns (Zheng & Greenberg, 2018). Amongst the variety of parameters and techniques, Wirzberger et al. (2018) recorded participants' mean normalized skin conductance response, indicating changes in the sympathetic nervous system (Chen, Zhou, & Yu, 2018), and heart rate, accompanying cognitive processing demands (Kennedy & Scholey, 2000). These measures have already proven sensitivity in related research (Chen et al., 2018) and point towards higher demands on cognitive resources by increasing values. However, they provide only an overall evaluation of the prevalent cognitive load level, without specifying different facets. A sophisticated conceptual approach to obtain information on individual cognitive load facets on a neural level was postulated by Whelan (2007). It aligns to existing evidence from functional neuroimaging literature that builds around the measurement of peaks in the blood oxygen level due to neuronal activity. Based on this rationale, he suggests that extraneous cognitive load would correspond in particular to activity in brain regions responsible for sensory processing, such as the posterior parietal association cortex, Broca's area, and Wernicke's area. By contrast, the intrinsic cognitive load component should be associated with activity in brain regions involved in maintaining and manipulating the attentional focus, in particular, the dorsolateral prefrontal cortex. Finally, germane cognitive load is assumed to hold connections to activity in brain regions related to motivation, as highly motivated learners are more likely to devote available cognitive resources solely to processes and strategies of schema acquisition. Corresponding brain regions, in this case, involve the superior frontal sulcus and the intraparietal sulcus. Although this approach offers high explanatory potential, so far it has not been explicitly tested yet due to the high methodological effort and inherent task-related constraints.

In terms of behavioral responses, duration-based parameters from speech signals have proven sensitivity to changing levels of cognitive load (Chen et al., 2016). They can be classified as behavioral, since they show inherent characteristics such as disfluency, articulation rate, content quality, the number of syllables, and the number and duration of pauses regardless of the meaning of the utterance. Existing evidence indicates that increasing levels of cognitive load result in a slower speech tempo as well as more and longer pauses within the speech flow due to necessary planning processes (e.g., Müller, Großmann-Hutter, Jameson, Rummer, & Wittig, 2001). Contrary to existing work that applied speech parameters to capture fixed task demands across shorter time spans (e.g., Yap, 2012), Wirzberger et al. (2018) inspected the phoneme-based features articulation rate, number of silent pauses, and duration of silent pauses with references to task-inherent processual changes during schema acquisition. Related work extended the focus by additional parameters and further applied the acoustic-prosodic features loudness and pitch, and the voice quality features vocal fold frequency and voice amplitude (Herms, 2018; Herms, Wirzberger, Eibl, & Rey, 2018). A comparison of discrete classes of low, medium, and high levels of cognitive load showed statistically significant differences for articulation rate, pause duration, pitch, and voice quality features. The latter indicate less rough or hoarse characteristics of the speech signal with increasing levels of cognitive load.

Considering the outlined characteristics, Korbach et al. (2017) already demonstrated the benefit of combining measures related to behavior and secondary task performance to achieve a continuous cognitive load assessment. Further accounting for the fact that a single measure alone is not sufficient to obtain the underlying pattern of cognitive resource investment in learning situations, Chen et al. (2018) emphasize an even more comprehensive approach on cognitive load assessment. They introduce a multimodal framework that fuses a variety of cognitive load-indicating channels of continuous learner-related information, for instance, physiological signals from skin conductance and electroencephalography and behavioral data streams from speech, gaze patterns, and mouse and keyboard interactions.

1.4 Research focus

The current thesis addresses ongoing controversies in instructional cognitive load research by examining the processes and mechanisms of the interplay between the outlined cognitive load facets. Corresponding to a process-related reconceptualization of the existing three-component model that takes into account temporal changes in resources related to schema acquisition across the task, it applies a combination of continuous approaches for cognitive load assessment. The resulting evidence should provide a foundation for the development of adaptive instructional procedures with learner-aligned instructional support. Considering this background, fading instructional guidance then can tie in with both the level of expertise, reflected in current performance, but also the level of invested cognitive resources, detectable by capturing learners' cognitive load. The inclusion of context-related features has further indicative value, for instance, due to a facilitative use of interruptions to keep learners involved in the task.

2 Experimental studies

On methodological accounts, the novelty of this thesis consists in applying experimental task frameworks from basic cognitive research to inspect the interplay of cognitive load facets in a controlled, internally valid manner. Due to the demonstrated impact of prior knowledge on task performance and cognitive resource demands (e.g., Chen et al., 2017; Rey & Buchwald, 2011; Rey & Fischer, 2013; Seufert, 2018), tasks with no or commonly shared prior knowledge were chosen to keep the arising influences at a constant level. A related joint characteristic of the reported studies constitutes the a priori determination of task complexity according to the number of interacting elements of information (Beckmann, 2010; Chen et al., 2017; Ngu et al., 2018; Sweller & Chandler, 1994). As a particular focus was put on the inspection of the learning process, a set of continuous measures was applied. Whereas data in the first and second study were collected in group-based settings, the third study used individual testing sessions due to the nature of the recorded measures. *Table 2* provides an overview of sample characteristics across the included studies and *Table 3* outlines details of the underlying research designs.

Table 2

Sample characteristics of the experimental studies reported in the included articles

Study	**N**	**M** [a]	**SD** [a]	**Range** [a]	**Gender** [b]
1	96	24.35	4.81	18-48	79.17
2	116	23.25	4.34	18-44	80.17
3	123	22.67	3.55	18-34	76.42

Note. [a] Age in years. [b] Percentage of female participants.

2.1 Methods

A main characteristic of the first experimental study (Wirzberger et al., 2016) comprised the investigation of the previously discussed facets of cognitive load in a joint task framework. The adopted paradigm of working memory updating (Ecker, Oberauer, Lewandowsky, & Chee, 2010) can be regarded as condensation of learning-relevant working memory processes, as content-related changes need to be represented correctly over time. In such tasks, an initially presented input undergoes several steps of updating, which involve processes of retrieval, transformation, and substitution that are reflected as well in a more implicit manner in the concluding recall of the final state. Aligned to Ecker et al. (2010), the presently employed task was formed of letter sets and accompanying alphabetic transformations over three (practice phase) or six (test phase) steps. Participants received a new set of letters at the outset of a trial that had to be incremented at one of the positions within each updating step and memorized afterward. As displayed in *Table 3*, facets of intrinsic, extraneous, and germane cognitive load were addressed by controlled, task-inherent variations (Beckmann, 2010) in a 3 x 2 x 2 within-subjects design: Firstly, the number of letters to be memorized simultaneously determined task complexity and resulted from adding or removing one letter around the intermediate difficulty of three letters (Ecker et al., 2010). Secondly, an increased distance between presented letters aligned to the split attention effect (Ginns, 2006; Schroeder & Cenkci, 2018) as means of inappropriate instructional presentation. Thirdly, the repetition of letter sets from a previous training sequence enabled the use of already existing task-related schemata. As individual aptitude variables are known to play an important role in such tasks, the standardized psychological inventory d2-R (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010) provided insights into participants' concentration abilities. Due to the recording of error rates (corrected for inherited errors) and reaction times in both update steps and final recall as dependent variables, a verifying feedback (Shute, 2008) on the percentage of correct responses could be provided after completing the task.

Table 3

Study designs of the experimental studies reported in the included articles

Study	Trials	Session duration	Material	ICL	ECL	GCL
1	6 [a] / 24 [b]	45 min	d2-R Letter sequences	Number of letters (2/3/4)	Increased distance of letters [c]	Repetition of letter sets [c]
2	64 [d]	60 min	OSPAN [e] SSPAN [e] Symbol sequences	Number of symbols (2/3)	Interrupting visual search task	Performance efficiency (cr/rt*1000)
3	64 [d] / 60 [f]	60 min	OSPAN Symbol sequences [g]	Number of symbols (3/4)	Embedded secondary task	Performance efficiency (cr/rt*1000)

Notes. ICL = intrinsic cognitive load, ECL = extraneous cognitive load, GCL = germane cognitive load. [a] Practice phase, [b] Test phase, [c] With vs. without, [d] Learning task, [e] Short versions, [f] Transfer task, [g] An additional classification task applied symbol sequences with distortions to assess transfer demands.

In the subsequent second experimental study (Wirzberger, Esmaeili Bijarsari, et al., 2017), the focus was shifted towards the processual nature of schema acquisition. The task further addressed the issue of potentially occurring interferences of prior knowledge from the previous letter stimuli and used more abstract material to inspect the temporal interplay of cognitive load facets. In more detail, participants had to learn abstract geometric symbol combinations via trial and error by verifying feedback (Shute, 2008) that informed about the correctness of the response and the correct response in terms of errors. In line with the first study, the number of symbols in a defined order that formed a combination represented task complexity as between-subjects independent variable. Under the assumption that the prevalence of distracting tasks with competing goals (Gerjets et al., 2003) represents a common situational constraint in computer-based learning environments, interruptions were induced at five defined stages over the task as further within-subjects independent variable. The emerging effects on performance should further hint on the underlying progress in schema acquisition. Following Wickens

(2002), a resource-demanding perceptual task would be able to cause substantial interference with a cognitive task that involves storage and/or transformation processes in working memory. On this account, the interrupting task itself adopted a visual search paradigm with a sufficient number of geometric symbols similar to the learning task (Trick, 2008), as similarity (Gillie & Broadbent, 1989) and an appropriate task duration (Monk, Trafton, & Boehm-Davis, 2008) should ensure the interrupting potential. To assess the investment of cognitive resources related to schema acquisition in the resulting 2 x 5-factorial mixed design, performance efficiency computed from reaction times and correct responses (Hoffman & Schraw, 2010), was inspected as dependent variable. Participants working memory span (Unsworth, Heitz, Schrock, & Engle, 2005) and perceived mental load and mental effort (Krell, 2015) were obtained before the learning task, whereas the amount of recalled symbol combinations was recorded afterward.

The third experimental study (Wirzberger et al., 2018) extended the abstract visual-motor symbol sequence learning task used in the second study by an embedded auditory-verbal secondary task to enable a closer monitoring of the investment of cognitive resources related to schema acquisition. Distinct input and output modalities were chosen to ensure the occurrence of resource interference merely at a cognitive stage due to the simultaneous processing of task requirements (Wickens, 2002, 2008). The constant interchange of both tasks over time was inspired by the procedure of automated complex working memory span tasks (Redick et al., 2012; Unsworth et al., 2005), which are characterized by the alternating sequence of distractor and target tasks. While the primary task slightly adjusted the task paradigm by Wirzberger, Esmaeili Bijarsari, et al. (2017) in terms of both number and presentation of symbols, the secondary task required participants to memorize and recall a spoken five-digit sequence from start to finish of each trial. Again, task complexity in the primary task varied according to the interrelated number of symbols. In addition to performance parameters from primary and secondary tasks, inspected via combined efficiency measures (Hoffman & Schraw, 2010), cognitive load-related parameters from prosodic speech features and physiological parameters were recorded. Particularly the inspection of varying levels of cognitive load in speech-related characteristics, such as the number and duration of pauses and the articulation rate, comprises a novel and innovative solution in cognitive load research (Herms et al., 2018). Accompanied by more established physiological markers of skin conductance response and heart rate, the study provided an elaborated pattern of multimodal indicators for cognitive load. Beyond a subsequent recall of memorized symbol combinations, a specifically designed transfer task aimed at obtaining the robustness of the acquired schemata. Based on the set of previously learned symbol combinations, it required participants to categorize displayed symbol combinations in terms of their correctness. The task operated on a 2 x 5 factorial mixed design that aligned with the aforementioned between-subjects variation of task complexity due to the number of symbols. In addition, defined levels of distortion of the presented symbols induced increased transfer demands that were inspected in terms of errors (corrected for inherited errors) and reaction times. Again, participants' individual working memory capacity was controlled for by completing a working memory span task (Unsworth et al., 2005) before the learning task.

2.2 Results

Responding to the request of Martin (2018) to apply more complex statistical models to represent the factor time in cognitive load assessment, data analysis across the included experimental studies is characterized by advanced statistical approaches (see *Table 4*). The resulting continuous inspection of learner states further corresponds to Leppink and Van Merrienboër (2015), who advise against the aggregation of repeated measures due to the resulting loss in informative value about individual task-related progressions.

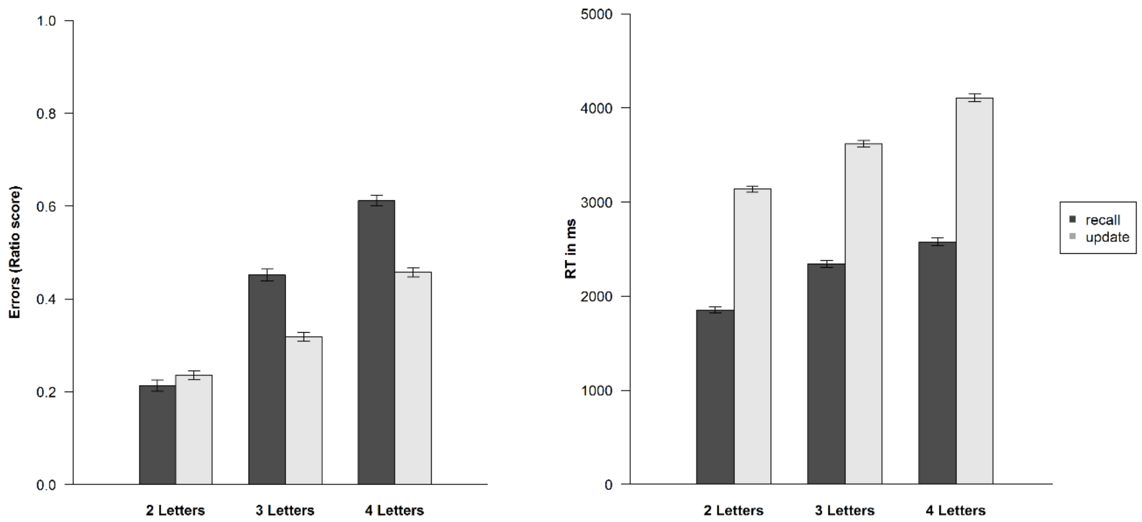

Figure 1. Complexity-related differences in reaction times and errors in both update and final recall stages. Error bars indicate standard errors.

In the first experimental study, results displayed a constant increase in both error rates and reaction times with increasing complexity, as shown in *Figure 1*. The visual impression suggests that participants reacted slower during the update steps but made more errors during the final recall. A significant increase in reaction times, but not error rates, with increased distance between stimuli resulted and the pattern of effects further indicated a benefit of repeated letter sets in terms of error rates and reaction times. In addition, significant two- and three-way-interactions between the examined facets showed up. Besides the beneficial influence of higher individual levels of concentration, task-related progress also fostered an overall increase in performance.

Table 4

Characteristics of data analysis in the experimental studies reported in the included articles

Study	**Participants included**	**Statistical approach**	**Main outcomes**
1	92	Linear mixed effects model [a]	Significant main effects and interactions Significant influence of aptitude and control variables
2	113	Conditional growth curve model [a] ANCOVA [a, b] t-Test	Superiority of logarithmic progression model Differences in interruption effects across conditions
3	103 [c]	Conditional growth curve model [a] Time series regression [d] ANOVA t-Test	Increasing performance under decreasing levels of cognitive load Robust transfer performance even under increased task demands

Notes. [a] Accounting for interindividual variance by inclusion of random intercept, [b] Based on linear mixed effects model, [c] Further exclusions required for secondary task performance (n = 102), speech parameters (n = 102), and physiological parameters (n = 101), [d] Accounting for interindividual variance by normalization on individual baseline.

Inspired by temporal models of learning and skill acquisition (Anderson, 1982; Ebbinghaus, 1964; Fitts & Posner, 1967), plausible linear, quadratic and logarithmic progressions were compared statistically in the second experimental study. The obtained results revealed a nonlinear increasing development of performance efficiency over time that differed between both conditions of task complexity, with superiority for the logarithmic model. In addition, condition differences arose with respect to the impairing influence of the induced interruptions, as a loss in performance was more obvious in the easy task condition. No differences in performance efficiency between easy and difficult task conditions resulted in the interrupting task. In addition, the amount of correctly recalled symbol combinations after completing the learning task was comparable between both conditions.

The pattern of evidence in the third experimental study indicated an increase in performance efficiency over time for both primary and secondary tasks. Performance benefits for the easy task condition were also supported by more correctly recalled symbol combinations after the learning task. In line with evidence from the second study, analyses of continuously recorded parameters were based on logarithmic progression models. Speech-related parameters pointed towards reduced levels of cognitive load with increasing task progress, as the number and mean duration of pauses decreased, and articulation rates increased. In a similar way, physiological parameters displayed decreasing progressions and furthermore showed a repetitive seasonal pattern across subtasks with increases in secondary task-related steps. Results obtained from the transfer task hint on robust and stable performance even under enhanced task demands due to distorted symbols. Reaction times were significantly increased, but generally faster under difficult task conditions.

2.3 Implications

Taken together, the pattern of evidence arising from the outlined experimental studies supports a process-related refinement of the theoretical framework of cognitive load, considering ICL and ECL on a structural and GCL on a processual level. Such also relates to the recently introduced reformulation of GCL as resources dealing with relevant aspects of a learning task (ICL) contrary to independent additivity (Sweller, 2018). On practical accounts, the obtained results provide the basis for developing multimodal models of cognitive load progression as algorithmic base for adaptive instructional support according to learners' individual cognitive resource supply.

Examining the results of the first experimental study in more detail, the prevalence of significant interactions supports existing doubts on the assumption of a purely additive relationship between the described cognitive load facets (Kalyuga, 2011; Sweller, 2018). In extension, the overall improvement in performance across the task emphasizes both the declarative and procedural nature of task-related schemata (Gagnè & Dick, 1983). The significant influence of increased concentration resources aligns with existing evidence on individual aptitude variables (e.g., Wirzberger & Rey, 2018) and emphasizes the importance of considering individual cognitive abilities in the context of learning. Increased demands due to enlarged spatial distance could be compensated by extended reaction times, whereas the overarching effect of increasing task complexity affected both measures without compensation. In addition, the study confirms the prevalence of the previously outlined processes involved in working memory updating (Ecker et al, 2010). These processes are directly reflected in increased reaction times during updating steps compared to the more indirect reflection in error scores during the final recall.

Summarizing the evidence obtained from the second experimental study, the superior logarithmic curve progression receives support from the well-established learning curve from Ebbinghaus (1964). On this account, the assumption that invested cognitive resources decrease at the same pace as learning performance increases receives reasonable corroboration but needs to be explored in more detail. Furthermore, condition effects in resumption performance relate to prior studies on volitional protection against competing goals (Gerjets et al., 2003; Scheiter, Gerjets, & Heise, 2014). Evidence shows that higher levels of task difficulty can shield against distractions from task-irrelevant information. In addition, such desirable difficulties could force people to apply alternative task-related strategies over time, for instance a more time-efficient heuristic encoding procedure that focuses just on a minor set amongst all offered cues.

The demonstrated increasing performance in both primary and secondary task in the third experimental study, accompanied by decreasing progressions in speech-related and physiological parameters, supports the assumption of decreasing levels of cognitive load due to increasing schema acquisition. As already outlined by Kraiger et al. (1993), such pattern hints on dynamics related to primary task automation, as free resources from this task can be increasingly devoted to deal with secondary task requirements. The evident seasonal pattern raises the conclusion that the auditory-verbal modality combination puts higher demands on learners' cognitive resources compared to the visual-motor modality combination. Since one potential explanation refers to the persisting dominance of the visual modality in many task domains, in instructional scenarios predominantly visual cues might be chosen for additional support. Similar to the results of the second experimental study, task performance in the transfer task again suggests a higher investment of cognitive resources with increasing task complexity. Tying in with evidence on desirable difficulties (Bjork & Bjork, 2011) and the zone of proximal development (Vygostki, 1963), to foster optimal learning performance, adaptive task procedures should provide constant challenges to keep learners involved in investing cognitive resources to achieve a robust and stable performance. On a methodological level, the correspondence between the applied measures particularly underlines the benefit to explore the potential of speech-related cognitive load indicators in multimodal learning environments.

2.4 Limitations

Although the first experimental study indicates statistically significant interactions between the addressed facets, these might have resulted due to interference in the experimental manipulation. In particular, the induced spatial distance highly depended on the number of presented letters, as a closer spatial proximity was required if there were more letters on the screen. Moreover, familiarity with the Latin alphabet could not be fully controlled and usually differs even amongst native speakers, which resulted in task-inherent benefits for participants with higher fluency and exposure.

From the obtained measures in the second experimental study, conclusions on the underlying progression of cognitive load result solely by the inversion of the resulting performance curves. Following Martin (2018), combined scores from invested time and obtained performance lack controllability, as participants could have reached equivalent levels of efficiency with different amounts of invested resources or achieved performance. Thus, a continuous monitoring of related resource demands is lacking as well as the further inclusion of motivational aspects. The latter could also have influenced how participants dealt with the task across different conditions of complexity. Since the task required participants to memorize only a few symbol combinations, participants facing increased complexity might have benefitted more from extended time frames due to presentation characteristics. Comparable to the previous study, the group-based testing sessions always involve the prevalence of peer-pressure in task-related timing.

The latter aspect was addressed in the third experimental study due to the use of individual testing sessions, as well as the alignment in terms of stimuli presentation between conditions. However, differences in symbol complexity might have resulted from varying visual characteristics of the used symbols, like the salient edges of a star. These could have fostered benefits in terms of the retentivity of certain symbol combinations. Moreover, task order ambiguities could have occurred, since the secondary task was presented first and interleaved by the primary task. Inspecting the obtained measures more closely, progressions in physiological parameters may hint on the prevalence of an orientation response at the outset of the task, followed by the adjustment with increasing task progress. However, even after the first ten trials, a recognizable progression persists that hints on a modified pattern of cognitive resource investment.

2.5 Future work

The first experimental study mainly indicated a reduction of task complexity, the use of more abstract material without the reliance on previous knowledge, and the focus on the more processual characteristic of schema acquisition. These aspects were addressed in the second experimental study that demonstrated the necessity to continuously monitor the task-related investment of cognitive resources. In addition, the question persisted how robust the acquired schemata would be under conditions of enhanced transfer demands. While the third experimental study could seize the outlined suggestions by applying a multi-method approach to cognitive load assessment, and a subsequent classification task for assessing transfer performance, it still requires the inclusion of motivational characteristics in future studies. Further valuable perspectives could arise from the use of additional parameters such as gaze movements or mouse interaction patterns as well as the transfer into more applied task domains. Moreover, exploring the use of different secondary task paradigms that involve incompatible modality-content matchings and combine auditory-motor and/or visual-verbal channels could extend the informative value in terms of the robustness of the observed patterns. On the level of data analysis, the application of more complex procedures for inspecting stages across the underlying cognitive load progression, such as hidden Markov models (e.g., Visser, Raijmakers, & Molenaar, 2002), could further increase the predictive scope of the obtained insights. For purposes of classifying and interpreting multimodal cognitive load-related signals, the additional use of machine learning approaches can be of value, as demonstrated by Herms (2018).

3 Cognitive modeling

To strengthen and extend evidence obtained from the second experimental study (Wirzberger, Esmaeili Bijarsari, et al., 2017), a cognitive modeling approach using the cognitive architecture ACT-R (Adaptive Control of Thought – Rational; Anderson, 2007) constitutes a further methodological building block of the current thesis. The related purposes are twofold: Firstly, potential explanations for the unexpected effect of induced interruptions should be explored. Secondly, further insights into the cognitive processes behind the postulated facets of cognitive load should be obtained on a neural level by region-of-interest (ROI) predictions (Borst & Anderson, 2017) based on evidence from functional magnetic resonance imaging (fMRI). A pilot version of this model for the easy task condition was reported in Wirzberger, Rey, and Krems (2017) and has been expanded since to both conditions.

As a great strength, a cognitive modeling approach requires a precise formalization of human cognition, since it raises the need to decompose steps within the given task and related cognitive actions. Based upon the close compatibility with vested psychological evidence on human information processing, such offers the opportunity to derive well-founded explanations on behavioral phenomena. The idea of building computational models to explain cognitive phenomena has already been discussed by Wegener (1967), who outlined the indicative value of an electronic simulation of mental processes for deriving and validating the related hypotheses. Under the presumption of an existing analogy between model and psychological processes, it allows studying mental functions under conditions that would be difficult or even impossible to realize in human experiments. The constant interchange between experiment and simulation permits to verify and rethink given hypotheses on behavioral patterns and underlying cognitive strategies, and thus opens up the “black box”. In the context of the Cognitive Load Theory, there have been cognitive modeling accounts as well. Sweller (1988) used the production system PRISM (Langley & Neches, 1981) to explain cognitive load effects in problem-solving. He compared means-end and nonspecific goal strategies by determining the number of statements in working memory, the number of productions to fire, the number of conditions in productions to be matched, and the number of cycles to be executed. His conclusions indicate that the conventional means-end problem-solving strategy puts higher demands on cognitive resources and not necessarily fosters schema acquisition.

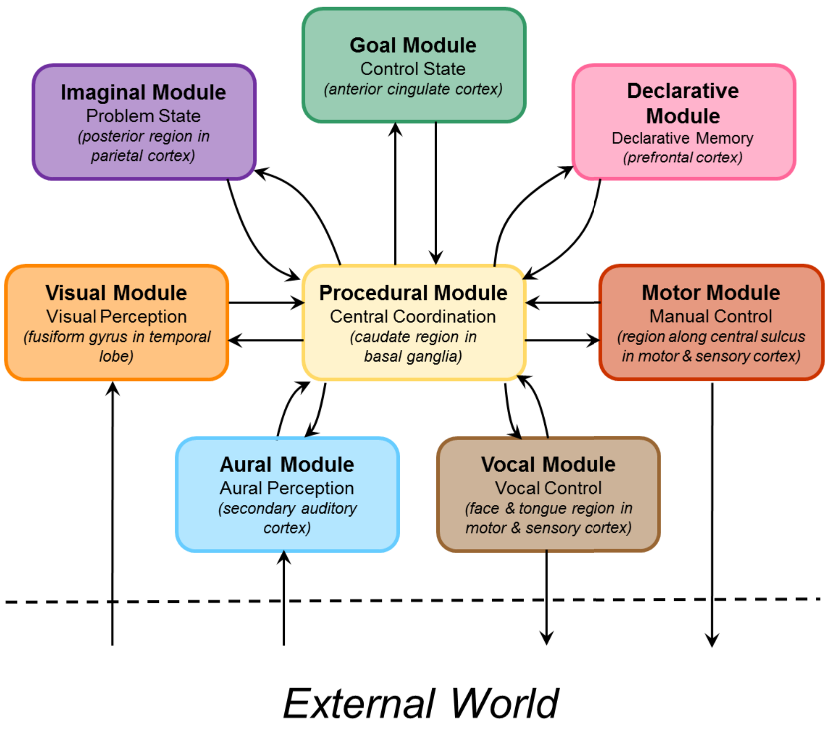

Figure 2. Overview of ACT-R core modules with corresponding brain regions. Based on Borst & Anderson (2015) and Anderson (2007).

Constituting a more prevalent and broadly used production-based approach, ACT-R is particularly characterized by its modular brain-inspired structure that is illustrated in *Figure 2*. The outlined modules represent goal planning (goal module), declarative memory (declarative module), intermediate problem states (imaginal module), action coordination (procedural module), the handling of visual and auditory inputs (visual and aural module), and motor and vocal outputs (motor and vocal module). Borst, Nijboer, Taatgen, van Rijn, and Anderson (2015) validated the mapping between these modules and corresponding ROIs in the human brain by fMRI data. For instance, when a model presses a button, increased activity in the motor module corresponds to activity in the motor cortex devoted to the representation of the hand. Whereas processes in different modules can be executed in parallel, known bottlenecks in information processing are represented by a limited capacitiy of a single information element per module at the same time (e.g., Borst, Taatgen, & van Rijn, 2010; Byrne & Anderson, 2001; Salvucci & Taatgen, 2008).

ACT-R relies on both symbolic and subsymbolic characteristics. The former involve the representation of declarative knowledge via so-called chunks of information and the interaction of defined modules through production rules. The latter constitute activation levels in declarative memory and utility of production rules. Chunks from declarative memory are retrieved based on their level of activation, which is calculated from the history and context of use and has to exceed a defined threshold to be eligible for selection. The full equation[1] for each chunk i involves the components displayed in *Equation 1:*

$$A_i = B_i + \sum_k \sum_j W_{kj} S_{ji} + \sum_l P M_{li} + \varepsilon. \tag{1}$$

The recency and frequency of use of the chunk i is reflected by the base-level activation B_i, W_{kj} represents the amount of activation from source j in buffer k, S_{ji} is the strength of association from source j to chunk i. W_{kj} and S_{ji} are summed over all buffers that provide spreading activation and all chunks in the slot of the chunks in buffer k. P reflects the amount of weighting given to the similarity in slot l and M_{li} represents the similarity between the value l in the retrieval specification and the value in the corresponding slot of chunk i. M_{li} is summed over the slot values of the retrieval specification. The value of ε represents noise, which is composed from an instantaneous component that is computed at the time of a retrieval request, and a permanent component that is associated with each chunk. Base-level activation is calculated as shown in *Equation 2*:

$$B_i = ln\left(\sum_{j=1}^{n} t_j^{-d}\right). \tag{2}$$

[1] Equations on chunk activation, base-level activation and utility relate to content described in the ACT-R reference manual and the tutorial units, available via http://act-r.psy.cmu.edu/software/.

It bases on the number of presentations *n* for the respective chunk *i*, the time t_j since the *jth* presentation, and a decay parameter *d*. Each time a chunk is presented, its base-level activation is increased, which decays as a power function of the time since that presentation. These decay effects are summed up and then transformed logarithmically.

Production rules consist of a condition part and an action part and are evaluated by the procedural module with regard to the content of the tested buffers. Based on the resulting pattern, a matching production rule is chosen, which triggers the related action. For instance, if the task is to react to a yellow number by key press, the visual module sees a yellow number, and the motor module is not in use, the action of pressing the key can be initiated. If more than one production rule fulfills the constraints, the selection of production i is informed by the subsymbolic cost-benefit mechanism of utility:

$$Probability(i) = \frac{e^{U_i\,/\,\sqrt{2s}}}{\sum_j e^{U_j\,/\,\sqrt{2s}}}. \tag{3}$$

It can be described as displayed in *Equation 3* by summing all productions *j* with expected utility values U_j that have matching conditions at the point of selection. Based on that, the production with the highest utility is chosen to fire.

3.1 Model concept

Each model run starts with an initial setting of the task goal, which is assumed to result from the previously read instruction. In the following, each learning trial builds upon three task-related steps, displayed in *Figure* 3. At first, the presented symbol is searched and encoded, which is repeated for the second symbol in the case of the difficult condition. This procedure stores an intermediate representation of all encoded visual content in the problem state, for instance, the input symbols ‘square – circle’ in the difficult condition. Next, the model attempts to retrieve the associated response symbol from declarative memory. In the second step, a response is selected from the provided opportunities on the screen, either according to the retrieved chunk or by random choice in case of no successful retrieval. The final step comprises the search for a visual feedback on the given response and, in the case of a false response, an update of the existing intermediate representation. The final information contains both the input and the correct response parts of the symbol combinations, such as ‘square – circle – square’ in case of the previous example. In the first trials, there is no sufficiently matching content or no content at all to retrieve, resulting in slower and less accurate responses. After being presented the input symbols several times and retrieving related content from declarative memory, the model performance gets increasingly faster and more accurate due to increasing chunk activation.

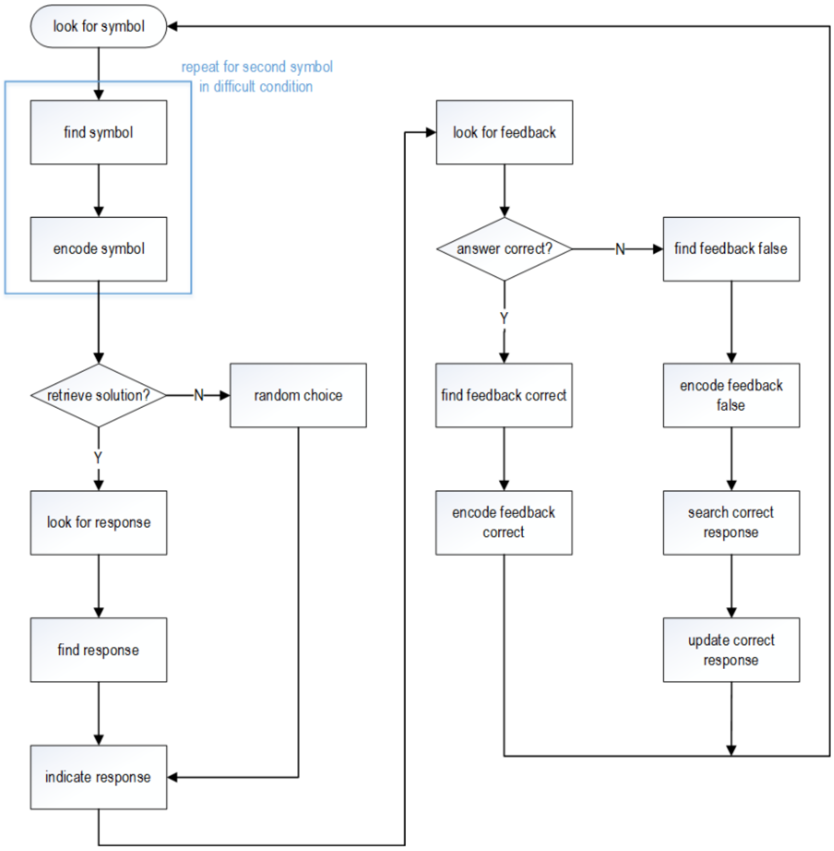

Figure 3. Outline of steps to perform in each the learning trial of the task. Adapted from Wirzberger, Rey, et al. (2017).

To account for the fact that humans sometimes retrieve related but ultimately wrong information from memory – in this case a wrong input-response association – ACT-R includes a partial matching mechanism. Based on initially defined similarities between chunks, a mismatch between request and actual retrieval is calculated. The higher the mismatch, the more the activity of the chunk is penalized (Lebiere, 1999). Increased interactivity between related elements of information (Sweller & Chandler, 1994; Sweller, 2010) is reflected in the spreading activation mechanism (Anderson, 2007) that distributes activation across chunks that share information elements. In the current task, spreading activation particularly effects the difficult task condition: Symbol combinations including the same input symbols, such as 'square – circle' and 'circle – square', obtain equal activation, independent of the correct symbol order.

The steps to be performed within the interrupting task are outlined in *Figure 4*. Following a goal change due to the bottom-up triggered saliency of the interrupting task, the task procedure involves the steps of searching, counting, and responding to the indicated target symbols. Using a color to indicate the task switch followed the model implemented by Wirzberger and

Russwinkel (2015) and represents the immediate attention to the related screen change. Tying in with evidence on pre-attentive and attentive processes in the visual module of ACT-R (Nyamsuren & Taatgen, 2013), the second visual-location request in the visual search is enhanced by additional information on stimulus color that relates to distinct characteristics of the presented symbols. In addition, counting was assumed to constitute a highly trained behavior that occurs almost automatically, thus a simple counting function was applied instead of intermediate retrievals after each counting step. After finishing the counting part, on each of the two response screens the model encodes the requested symbol and attempts to retrieves the potential answer. Again, due to the partial matching mechanism the possibility to retrieve a wrong answer persists.

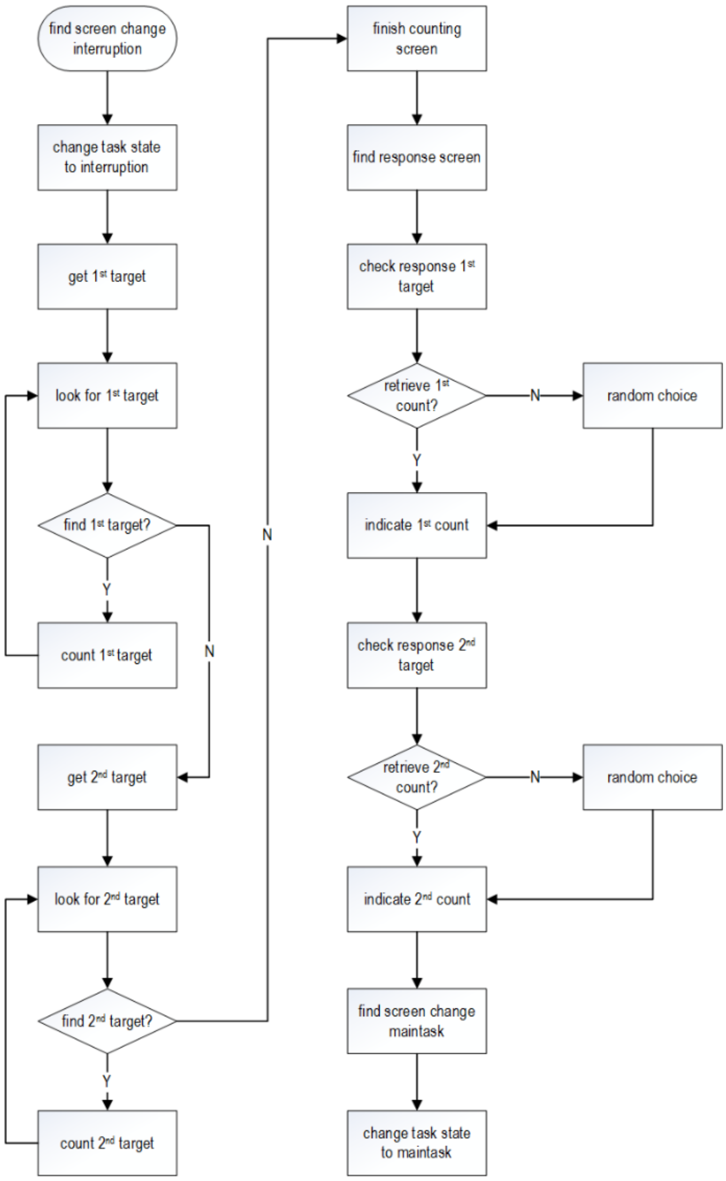

Figure 4. Outline of steps to perform in each occurrence of the interrupting task. Adapted from Wirzberger, Rey, et al. (2017).

Aligning to the available button selection on the screen, the model operates on an increased amount of visual-number finsts. Furthermore, due to the fixed order of the buttons on the screen, they can receive immediate attention without searching through all buttons from the top. When resuming the learning task, in line with Altmann and Trafton (2002) the model attempts to retrieve the previous task goal and thus restores its representation. Emerging interruption effects can be attributed to a decay in the activation of chunks related to the learning task that slows down subsequent retrieval requests (Borst, Taatgen, & van Rijn, 2010, 2015; Trafton, Altmann, Brock, & Minz, 2003). Across both tasks, more specific actions are regarded as more useful and thus receive a slightly higher utility, for instance, productions related to attending or encoding instead of searching around.

3.2 Model comparison

Parameter settings in the reported model are outlined in *Table 5* and correspond to the range of reported standard values (e.g., Anderson, Bothell, Lebiere, &, Matessa, 1998). In addition, the goal chunk related to the learning task received an initial base-level of 70 to account for the fact that participants received a comprehensive instruction on this task before. For the partial matching mechanism, similarities between symbols were set to -1, and for the spreading activation mechanism, content in the imaginal buffer was defined as source for spreading activation upon each retrieval from declarative memory. As already outlined, the increased amount of visual number finsts aligned to the button selection presented on the screen and received a value of 10.

Table 5

Parameter settings related to chunk activation and retrieval time

	:bll[a]	**:mp**[a]	**:mas**[a]	**:ans**[a]	**:rt**[b]	**:lf**[b]
Setting	0.5	0.401	1.7	0.2	0.11	3.8
Description	Base-level decay	Mismatch penalty	Maximum associative strength	Instantaneous noise	Retrieval threshold	Latency factor

Note. [a] Related to chunk activation. [b] Related to retrieval time (including retrieval failure).

Model data based on n = 100 model runs in each condition, since it was not the goal to create an exactly mapping model run for each human participant (n_{easy} = 55, $n_{\text{difficult}}$ = 58), but rather to obtain robust conclusions from the average model performance. In addition, a close behavioral mapping in terms of interruption performance was not the core focus of the model, thus in the following only comparative results regarding the symbol learning task will be reported in detail. However, it was ensured that no crucial differences between both conditions persisted for the interrupting task. Both human performance and the currently reported model meet this constraint.

When comparing human and model data, beyond a graphical inspection Schunn and Wallach (2005) recommend the combined consideration of numerical goodness-of-fit indices related to the relative trend magnitude and the deviation from the exact location. They suggest R^2 to assess the relative trend magnitude, as it directly refers to the accounted proportion of variance and indicates a better fit by higher values. It is particularly suited to evaluate models with strong correlations to human data. For obtaining the deviation from the exact location, the root mean squared scaled deviation (RMSSD) constitutes a sophisticated approach:

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{m_i - d_i}{s_i / \sqrt{n_i}}\right)^2}{k}} = \sqrt{\sum_{i=1}^{k}\frac{(m_i - d_i)^2 n_i}{k s_i^2}}. \quad (4)$$

As obvious from *Equation 4*, the RMSSD scales the deviation between the model mean m_i for each point i and the data mean d_i for each point i by the corresponding standard error of this mean from human data ($s_i / \sqrt{n_i}$). The latter is calculated by the standard deviation s_i for each data mean i and the number of data values n_i contributing to each data mean d_i, whereas k is the number of points i. On this account, the RMSSD provides a scale invariant measure to evaluate the model fit in units of the standard error, with lower values indicating a better fit.

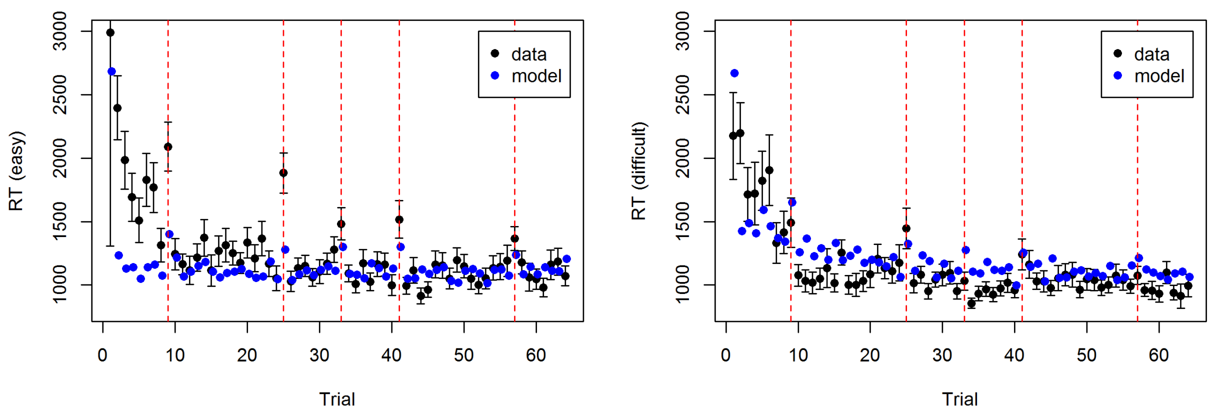

Figure 5. Reaction times for human data and model for the learning task in the easy and difficult condition (correct trials). Error bars indicate standard errors for human data. Red dashed lines indicate trials after an interruption.

In terms of reaction times, comparisons focused only on correctly solved trials. As obvious from *Figure 5*, interruption effects are observable in both conditions for human data, but still more distinctive in the easy task condition, as reported in Wirzberger, Esmaeili Bijarsari, et al. (2017). Standard errors are rather high for the first data point in the easy task condition, as only $n = 2$ observations fulfill the stated constraints. Besides the prevalence of interruption effects in both conditions, the visual inspection indicates that model data can map the initial decrease in reaction times in the difficult task condition, $RMSSD_{difficult} = 2.16$, $R^2_{difficult} = 0.58$. However, the model performs slightly slower than human participants during most of the trials. Apart from a subtler decrease in the beginning, the mapping fits quite well for later trials in the easy task condition, $RMSSD_{easy} = 1.67$, $R^2_{easy} = 0.52$.

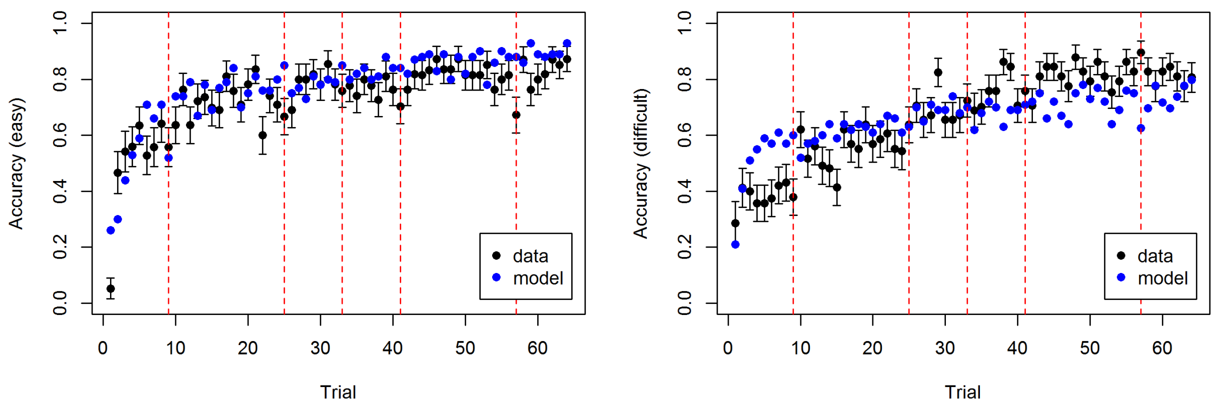

Figure 6. Accuracy for human data and model for the learning task in the easy and difficult condition. Error bars indicate standard errors for human data. Red dashed lines indicate trials after an interruption.

For accuracy, *Figure 6* indicates that the model can map the existing human behavior quite well in the easy task condition, $RMSSD_{easy} = 1.51$, $R^2_{easy} = 0.69$, although it achieves a higher performance in the end and shows a subtler reflection of interruption effects. The model in the difficult task condition learns slower compared to the easy task condition, but still faster than the human participants. However, apart from the nearly perfect location match in the last data points, it cannot fully map the final increase in the human data, $RMSSD_{difficult} = 2.07$, $R^2_{difficult} = 0.57$.

In addition, predefined ROI-predictions were generated (Borst & Anderson, 2017), based upon the previously amplified mapping of activity in ACT-R modules on defined brain regions. The underlying approach uses the recorded start and end times of module activity to simulate a signal comparable to the blood oxygenation level obtainable via fMRI, which shows peaks about 4-6 s after the occurrence of neuronal activity. In the first step, the activity of each inspected module is represented as 0-1 demand function and convolved afterward with the hemodynamic response function, displayed in *Figure 7*. As an example, related to the task of the current model, longer retrieval times due to lower levels of chunk activation would result in increased activity in the declarative module. Such patterns are expectable in early stages of the task, with increased task difficulty, or caused by interruption-related decay, and would be observable by higher peaks in the resulting simulated signal.

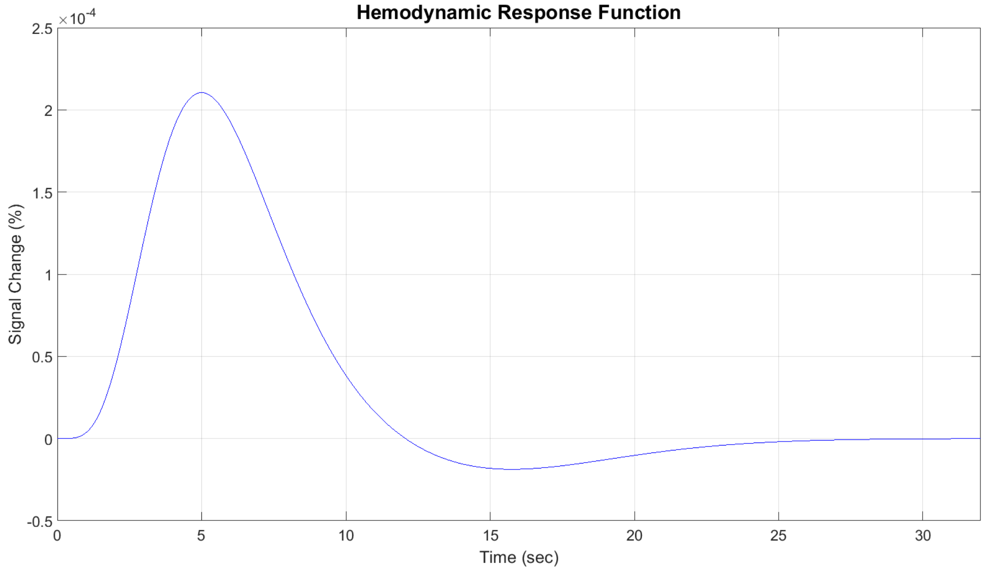

Figure 7. Hemodynamic response function (based on SPM). Adapted from Borst & Anderson (2017).

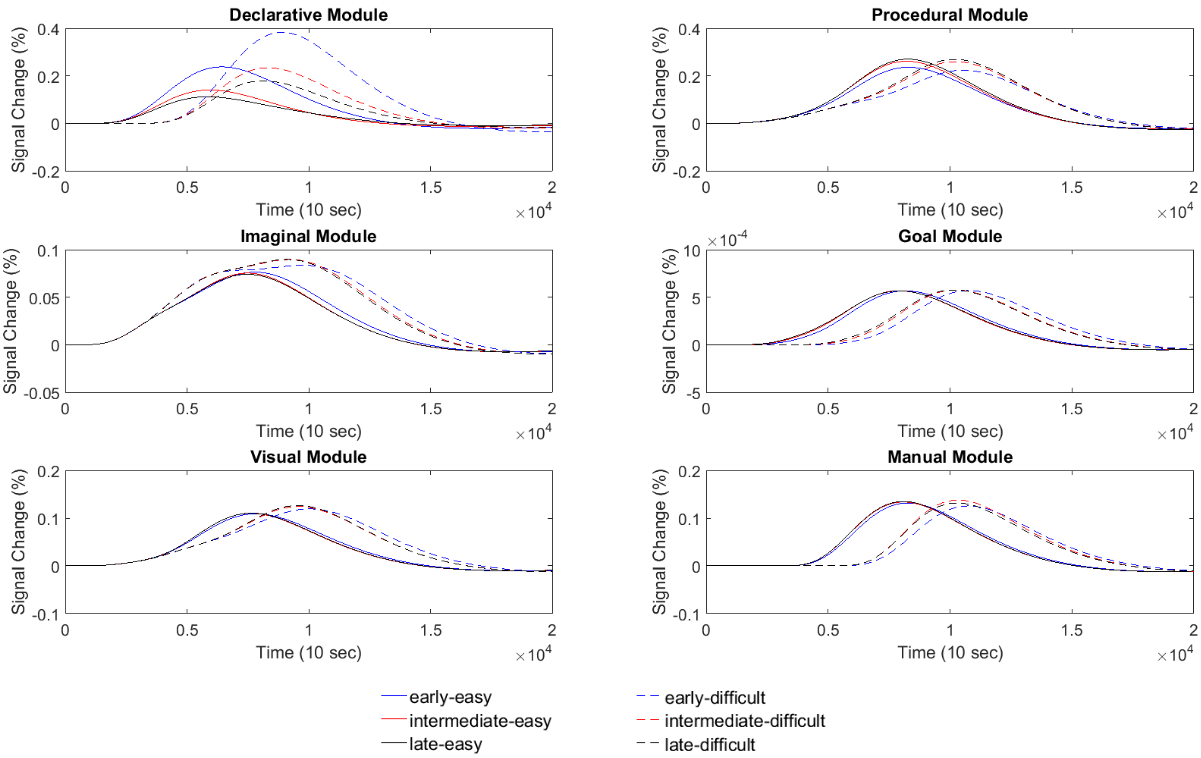

Figure 8. Module activity across different temporal stages of the symbol learning task (excluding resumption trials). Solid lines represent the easy task condition, dashed lines represent the difficult task condition. Blue lines represent trials in the early task stage (n = 20), red lines represent trials in the intermediate task stage (n = 19), and black lines represent trials in the late task stage (n = 20).

Prevalent changes in module activity due to task-inherent learning processes are displayed in *Figure 8*. The curves indicate a decrease in cognitive activity in later task stages in both conditions in the declarative module. The difficult task condition shows a higher level of activity across all stages, with a particularly distinctive peak across early task stages. Resulting activity in the imaginal module exerts a longer duration and shows a slightly increased level in the difficult condition. For the goal, procedural, visual and manual module levels of activity are rather comparable for both conditions, although the peaks occur later in the difficult condition.

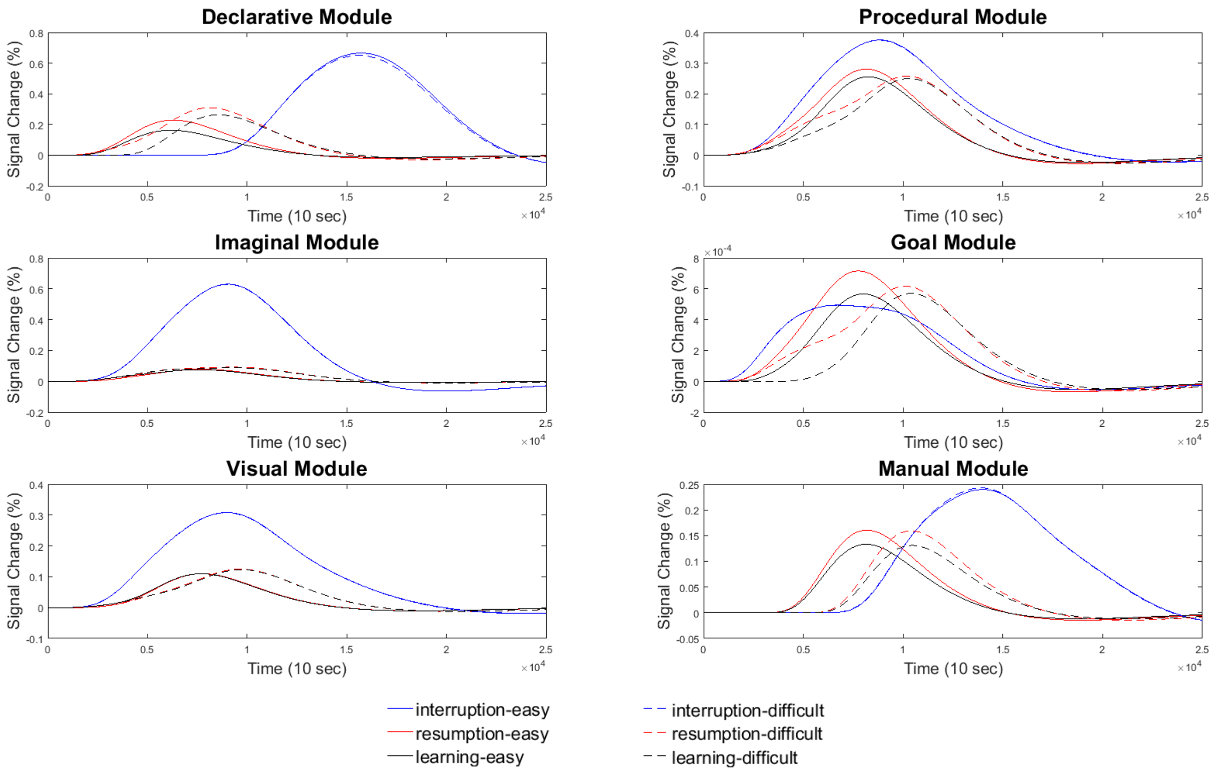

Figure 9. Module activity across interruption, resumption, and learning stages of the task. Solid lines represent the easy task condition, dashed lines represent the difficult task condition. Blue lines represent interruption trials (n = 5), red lines represent resumption trials (n = 5), and black lines represent learning trials (n = 59).

Comparisons between the interrupting task and the learning task are depicted in *Figure 9*. These include a separate visualization of the resumption phase, defined as the first trial that immediately follows the interrupting task. Across all inspected modules, activity levels in the interrupting task do not differ between both task conditions, since the solid and dashed blue lines overlap almost all the time. For the declarative, goal, procedural and manual module, a higher activity across resumption trials compared to the remainder of trials in the learning task results for both conditions. In addition, obvious differences between both task conditions show up during the resumption phase for the goal module and indicate higher levels of activity in the easy task condition. By contrast, no crucial differences between the resumption phase and regular learning trials result for the visual and imaginal modules. Apart from the goal module, the interrupting task always involves a higher level of activity, which peaks later in the declarative and manual modules.

3.3 Implications

The current model comprises a sophisticated approach to explore cognitive processes and mechanisms underlying changes in performance due to the inserted interruptions and task-related progress. In particular, the application of the spreading activation mechanism to map the theoretically introduced concept of element interactivity (Chen et al., 2017; Ngu et al., 2018; Sweller & Chandler, 1994; Sweller, 2010) offers the potential for deconstructing and formalizing effects of increased task complexity on a cognitive level. Inspecting model performance in the easy task condition in more detail, the obvious decrease in human performance in the final learning stage could potentially result from effects of boredom or fatigue. Modeling and explaining such effects would require a different model that also focuses on these effects (Gonzalez, Best, Healy, Kole, & Bourne Jr., 2011). In order to keep the current model focused and as simple as possible, this component was not included. For the difficult condition, model performance hints on an underlying shift in task-related strategies. Due to the small number of learned symbol combinations, over time people might have applied a more heuristic encoding strategy with focus on the first symbol, directly mapping task execution in the easy task condition. Explaining such strategy shift would result in a more complex model on the level of production rules and corresponding selection mechanisms. Taking this into account, the current modeling approach offers potential for future work, first by broadening the scope of the existing model and second by validating this model with new tasks. An additional benefit consists in explaining task order effects resulting from Wirzberger et al. (2018) with an additional ACT-R model that could build on the reported model. However, instead of dealing with an interrupting task, this model would face the constant requirement to simultaneously handle primary and secondary task procedures across both the visual and auditory modality.

As obvious from the ROI-analysis, the model needs to invest a higher amount of declarative memory resources upon each retrieval in the early task stage due to the lack of suitable chunks and lower levels of chunk activation. The smaller level of activity with increasing task progress emphasizes the prevalence of learning effects in both conditions, as existing content in declarative memory receives increasingly higher activation and thus can be retrieved faster and more accurate. In the difficult task condition, invested declarative resources are constantly higher across all stages, which by closer inspection relates to the increased influence of partial matching that penalizes chunk activation and extends retrieval times. It also corresponds well to the previously outlined conceptual approach by Whelan (2007). He attributed increased activity in the dorsolateral prefrontal cortex, a brain-region also connected to the declarative module, to higher levels of intrinsic cognitive load. The later peaks in activity in all modules in the difficult task condition potentially relate to attending and encoding an additional symbol, which, for instance, delays the onset of motor activity related to the response selection. Comparing activity patterns in both learning and interrupting tasks emphasizes the interrupting potential of the visual search task, since the activity in several modules clearly exceeds the activity during symbol learning. However, for both task conditions, task-related demands observable in the goal module are still higher in the learning task, hinting on more complex task-inherent control requirements. Similar activity patterns in both task conditions for the interrupting task reflect the absence of crucial differences between conditions and align to the pattern in human data reported by Wirzberger, Esmaeili Bijarsari, et al. (2017). The observable increased level of activity in the visual module during the interrupting task, which was supposed to trigger extraneous cognitive load, corresponds well to the reported activity in brain regions involved in sensory processing (Whelan, 2007). Finally, observable differences in goal activity during the resumption stage align well with predictions stated by the memory-for-goals model (Altmann & Trafton, 2002). They relate to the demand to rebuild the goal-representation of the learning task after each interruption, which also requires additional production rules, as reflected in increased procedural activity. Increased levels of resumption-related activity in the declarative module should arise from the decay of chunks related to the acquired symbol combinations. Finally, a reasonable explanation for the observable increase in motor-related activity in the resumption stage consists in the relocation of the mouse cursor from a different response screen.

4 Conclusions

The current thesis critically approached existing debates in cognitive load research related to the scope and interplay of distinct resource-demanding facets in instructional situations. Taken together, it emphasizes a process-related reconceptualization of the existing three-component model and underlines the importance of a combined inspection of different cognitive load measures. By extending the experimentally obtained behavioral results with a cognitive modeling approach, underlying cognitive processes and mechanisms could be inspected in more detail. The obtained insights further support the time-related reconsideration of the cognitive load facet framework, even on a neural level. With reference to applications in instructional situations, the resulting evidence can provide a vested foundation for the development of elaborated adaptive instructional procedures, both on the level of underlying algorithms and the design of instructional support. On this account, the research conducted within this thesis leverages a pathway to innovative approaches in the development of intelligent educational assistants.

References

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model, *Cognitive Science, 26*, 39–83. doi:10.1207/s15516709cog2601_2

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406. doi:10.1037/0033-295X.89.4.369

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA, USA: Harvard University Press.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY, USA: Oxford University Press.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341–380. doi: 10.1006/jmla.1997.2553

Atkinson, C. & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*, 82-90. Stable URL: http://www.jstor.org/stable/24922803

Ayres, P. (2018). Subjective measures of cognitive load. What can they reliably measure?. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 287-759). New York, NY, USA: Routledge.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559. doi:10.1126/science.1736359

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. doi: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.

Borst, J. P., & Anderson, J. R. (2015). Using the ACT-R cognitive architecture in combination with fMRI data. In *An introduction to model-based cognitive neuroscience* (pp. 339-352). New York, NY, USA: Springer.

Borst, J. P., & Anderson, J. R. (2017). A step-by-step tutorial on using the cognitive architecture ACT-R in combination with fMRI data. *Journal of Mathematical Psychology, 76*, 94-103. doi:10.1016/j.jmp.2016.05.005

Borst, J. P., Nijboer, M., Taatgen, N. A., van Rijn, H., & Anderson, J. (2015). Using data-driven model-brain mappings to constrain formal models of cognition. *PLoS ONE, 10,* e0119673. doi: 10.1371/journal.pone.0119673

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 363-382. doi:10.1037/a0018106

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2015). What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In *CHI '15. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2971-2980). New York, NY, USA: ACM Press. doi:10.1145/2702123.2702156

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2-Revision. Aufmerksamkeits- und Konzentrationstest*. Göttingen, Germany: Hogrefe.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*, 115-132. doi:10.1023/B:TRUC.0000021812.96911.c5

Brünken, R., Steinbacher, S., Plass, J., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 109-119. doi:10.1027/1618-3169.49.2.109

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review*, *108*, 847–869. doi: 10.1037/0033-295X.108.4.847

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust multimodal cognitive load measurement*. Bern, Switzerland: Springer International Publishing.

Chen, F., Zhou, J., & Yu, K. (2018). Multimodal and data-driven cognitive load measurement. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 3643-4043). New York, NY, USA: Routledge.

Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review, 29*, 393-405. doi:10.1007/s10648-016-9359-1

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current Directions in Psychological Science, 19*, 51–57. doi:10.1177/0963721409359277

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY, USA: Harper & Row.

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

D'Esposito, M., Onishi, K., Thompson, H., Robinson, K., Armstrong, C., & Grossman, M. (1996). Working memory impairments in multiple sclerosis: Evidence from a dual-task paradigm. *Neuropsychology*, *10*, 51. doi:10.1037/0894-4105.10.1.51

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Mineola, NY, USA: Dover Publications.

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. doi:10.1037/a0017891

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2017). Lernförderliche Gestaltung computerbasierter Instruktionen zur Roboterkonstruktion. In M. Eibl, & M. Gaedke (Eds.), *INFORMATIK 2017, Lecture Notes in Informatics (LNI)* (pp. 2279-2286). Bonn, Germany: Gesellschaft für Informatik. doi:10.18420/in2017_228

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Gagné, R. M., & Dick, W. (1983). Instructional psychology. *Annual Review of Psychology, 34*, 261-295. doi:10.1146/annurev.ps.34.020183.001401

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types?. *International Journal of Psychophysiology*, *83*, 269-275. doi:10.1016/j.ijpsycho.2011.09.023

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.

Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive?: A study of length, similarity, and complexity. *Psychological Research, 50*, 243-250. doi:10.1007/BF00309260

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511-525. doi:10.1016/j.learninstruc.2006.10.001

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema?: A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. doi:10.1016/j.neuropsychologia. 2013.11.010

Gonzalez, C., Best, B., Healy, A. F., Kole, J. A., & Bourne Jr., L. E. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research, 12*, 19-32. doi:10.1016/j.cogsys.2010.06.004

Herms, R. (2018). *Effective speech features for cognitive load assessment: Classification and regression* (Doctoral dissertation). Technische Universität Chemnitz, Germany.

Herms, R., Wirzberger, M., Eibl, M., & Rey, G. D. (2018). CoLoSS: Cognitive load corpus with speech and performance data from a symbol-digit dual-task. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, … T. Tokunaga (Eds.), *Proceedings of the 11th International Language Resources and Evaluation Conference* (LREC 2018) (pp. 4312-4317). Miyazaki, Japan: European Language Resources Association (ELRA).

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45*, 1-14. doi:10.1080/00461520903213618.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1-19. doi:10.1007/s10648-010-9150-7

Kalyuga, S., & Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371. doi: 10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kalyuga, S., & Plass, J. L. (2018). Cognitive load as a local characteristic of cognitive processes. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 1477-1884). New York, NY, USA: Routledge.

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831-852. doi:10.1007/s10648-015-9352-0

Kennedy, D., & Scholey, A. (2000). Glucose administration, heart rate and cognitive performance: Effects of increasing mental effort. *Psychopharmacology, 149*, 63-71. doi: 10.1007/s002139900335

Klepsch, M., Schmitz, F., & Seufert, T. (2017) Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*:1997. doi:10.3389/fpsyg.2017.01997

Korbach, A., Brünken, R., & Park, B. (2017). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review, 30*, 503-529. doi:10.1007/s10648-017-9404-8

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328. doi:0021-9010/93/S3.00

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

Langley, P., & Neches, R. (1981). *PRISM user's manual*. Technical report, Department of Psychology, Carnegie-Mellon University and Learning Research and Development Center, University of Pittsburgh, USA.

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft, 8*, 5–19. doi: 10.1007/BF03354932

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*, 1058-1072. doi:10.3758/s13428-013-0334-1

Leppink, J., Paas, F. G. W. C., Van Gog, T., Van der Vleuten, C. P. M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. doi:10.1016/j.learninstruc.2013.12.001

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

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie, Band C/I/5* (pp. 799-864). Göttingen, Germany: Hogrefe.

Martin, S. (2018). A critical analysis of the theoretical construction and empirical measurement of cognitive load. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 778-1155). New York, NY, USA: Routledge.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY, USA: Cambridge University Press.

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. doi:10.1037/a0014402

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., & Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In *International Conference on User Modeling* (pp. 24–33). New York, NY, USA: Springer Science + Business Media.

Ngu, B. H., Phan, H. P., Yeung, A. S., & Chung, S. F. (2018). Managing element interactivity in equation solving. *Educational Psychology Review, 30*, 255-272. doi:10.1007/s10648-016-9397-8

Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive Systems Research, 24*, 62-71. doi:10.1016/j.cogsys.2012.12.010

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

Paas, F., Tuovinen, J. E., Tabbers, H. und Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71. doi:10.1207/S15326985EP3801_8

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91. doi:10.1016/j.learninstruc.2006.02.004

Park, B., & Brünken, R. (2018). Secondary task as a measure of cognitive load. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 1901-2317). New York, NY, USA: Routledge.

Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load – an experimental dual-task study. *Applied Cognitive Psychology, 29*, 232-243. doi:10.1002/acp.3100

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164-171. doi:10.1027/1015-5759/a000123

Renkl, A., Gruber, H., Weber, S., Lerche, T., & Schweizer, K. (2003). Cognitive Load beim Lernen aus Lösungsbeispielen. *Zeitschrift für Pädagogische Psychologie, 17*, 93-101. doi:10.1024/1010-0652.17.2.93

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied*, 17, 33–48. doi:10.1037/a0022243

Rey, G. D., & Fischer, A. (2013). The expertise reversal effect concerning instructional explanations. *Instructional Science, 41*, 407–429. doi:10.1007/s11251-012-9237-2

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*, *115*, 101–130. doi: 10.1037/0033-295X.115.1.101

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. doi:10.3389/fpsyg.2014.00268

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

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review, 30*, 679-701. doi:10.1007/s10648-018-9435-9

Schunn, C. D., & Wallach, D. (2005). Evaluating goodness-of-fit in comparison of models to data. In W. Tack (Ed.), *Psychologie der Kognition: Reden und Vorträge anlässlich der Emeritierung von Werner Tack* (pp. 115-154). Saarbruecken, Germany: University of Saarland Press.

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. doi:10.1007/s10648-013-9242-2

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review, 24*, 116-129. doi:10.1016/j.edurev.2018.03.004

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153-189. doi:10.3102/0034654307313795

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285. doi:10.1207/s15516709cog1202_4

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

Sweller, J. (2018). The role of independent measures of cognitive load in cognitive load theory. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 155-267). New York, NY, USA: Routledge.

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. doi:10.1207/s1532690xci1203_1

Sweller, J., Van Merrienboër, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296. doi:1040-726X/98/0900-0251S15.00/0

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. doi: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. doi:10.3758/PP.70.5.743

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. doi:10.3758/BF03192720

Visser, Raijmakers, & Molenaar (2002). Fitting hidden Markov models to psychological data. *Scientific Programming, 10*, 185-199. doi:10.1155/2002/874560

Vygotski, L. S. (1963). Learning and mental development at school age In B. Simon & J. Simon (Eds.), *Educational psychology in the U.S.S.R.* (pp. 21-34). London, UK: Routledge & Kegan Paul.

Wegener, H. (1967). Die Simulation menschlicher Intelligenz – Leistungen. In W. Kroebel (Ed.) *Fortschritte der Kybernetik* (pp. 375-385). Munich-Vienna, Germany-Austria: R. Oldenbourg Verlag.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*, 1-12. doi:10.1016/j.edurev.2006.11.001

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159-177. doi:10.1080/1463922021012380 6

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*, 449-455. doi:10.1518/001872008X288394

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. doi:10.1016/j.compedu.2016.04.010

Wirzberger, M., Esmaeili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task. *Acta Psychologica, 179*, 30-41. doi:10.1016/j.actpsy.2017.07.001

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications*. doi: 10.1186/s41235-018-0138-z

Wirzberger, M., Rey, G. D., & Krems, J. F. (2017). Modeling cognitive load effects in an interrupted learning task: An ACT-R approach. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 3540-3545). Austin, TX, USA: Cognitive Science Society.

Wirzberger, M., & Rey, G. D. (2018). Attention please! Enhanced attention control abilities compensate for instructional impairments in multimedia learning. *Journal of Computers in Education, 5*, 243-257.doi:10.1007/s40692-018-0106-0

Wirzberger, M., & Russwinkel, N. (2015). Modeling interruption and resumption in a smartphone task: An ACT-R approach. *i-com, 14*, 147-154. doi:10.1515/icom-2015-0033

Yap, T. F. (2012). *Speech production under cognitive load: Effects and classification* (Doctoral dissertation). University of New South Wales, Sydney, Australia.

Zheng, R. (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice*. New York, NY, USA: Routledge.

Zheng, R. Z., & Greenberg, K. (2018). The boundary of different approaches in cognitive load measurement. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 1179-1472). New York, NY, USA: Routledge.

Article 1

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

Original article[1]:

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. doi: 10.1016/j.compedu.2016.04.010

Abstract

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 x 2 x 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.

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

[1] The following corrections were made to the original article: With reference to Table 1, within the note on Table 2, the notions of with and without were reversed, and in chapter 3.1, third paragraph, third sentence, the word "more" instead of "fewer" was inserted.

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 these 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 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 (Park, 2010; Brünken, Plass, & Moreno, 2010) as well as the separability of load facets (de

Jong, 2010; Kalyuga, 2011), 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 (η_p^2 = .37) the obtained results reveal a better performance without the requirement to store results of individual subtasks (η_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; 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. Seufert, 2003; Seufert & Brünken, 2006; Park, 2010). 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, e.g. 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, e.g. 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 (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_{age} = 24.35 years, SD_{age} = 4.81, 76 female), participated in the study. The sample split up into various disciplines of study, 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 x 2 x 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.

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 by 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.

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, those related to 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.

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

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 x 1050 px and a video refresh rate of 75 Hz. The task was implemented in *PXLab* (Irtel, 2007) with a timing precision better than 1 ms. After performing a written instruction including a detailed example, a set of six practice trials followed.

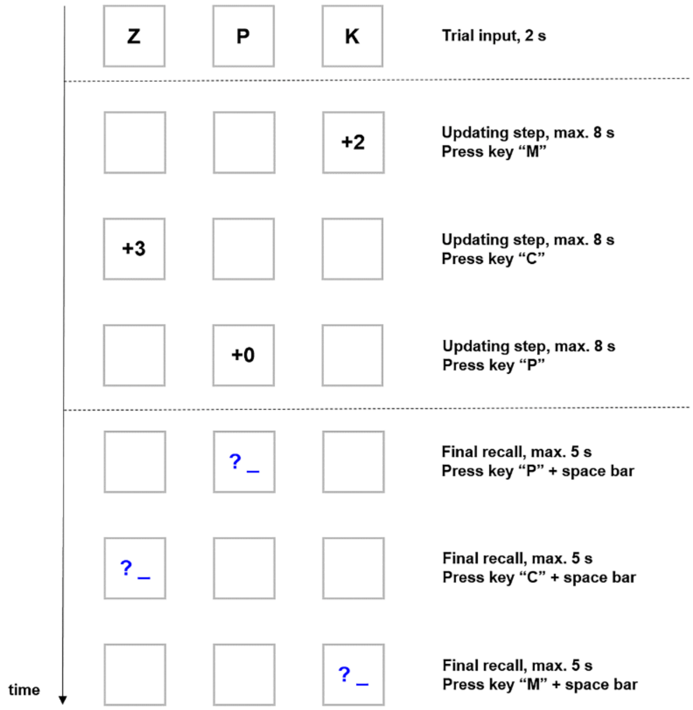

Figure 1. Sample practice trial sequence for the working memory updating task.

As illustrated in *Figure 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).

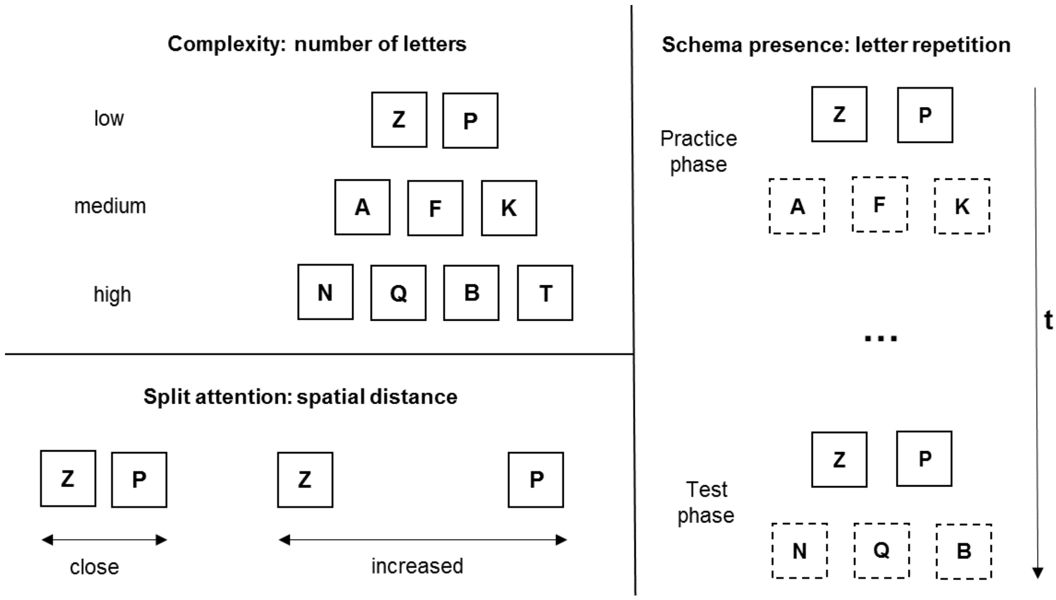

Figure 2. Experimental manipulations of complexity, split-attention and schema presence. Boxes with dashed lines indicate the lack of repetition in letter sets.

The 24 test trials entailed exactly the same procedure, except the inclusion of six instead of three update steps[2], 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). *Figure 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.

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

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

20 s for one line, participants received an experimenter command and proceeded with the following line until they completed the test sheet.

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 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 to 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

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 (RT_{update}), or via averaging the two to four observations within the final recall step (RT_{recall}). An analogous computation was performed for errors ($Errors_{recall}$, $Errors_{update}$), 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 $Errors_{recall}$, and six for $Errors_{update}$. On this account, values between zero and one resulted, indicating the actual amount of errors relative to the possible amount of errors. For $Errors_{recall}$

as well as $Errors_{update}$ 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[3].

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 ($N = 4\,019$), 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 level 1 and between-subjects' aptitude variables on level 2 (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 as well as between different models. For all dependent variables, models achieved similar fits on Akaike's information

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

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

3.1 Main effects

Analyses displayed in *Table 2* revealed constantly remarkable significant effects of complexity for RT_{recall}, $Errors_{recall}$, RT_{update} and $Errors_{update}$. 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.

Table 1

Descriptive values of dependent variables regarding main effects of independent variables

		RT_{recall} (ms)		**$Errors_{recall}$**		**RT_{update} (ms)**		**$Errors_{update}$**	
		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. RT_{recall} = reaction time during final recall, $Errors_{recall}$ = errors during final recall, RT_{update} = reaction time during updating steps, $Errors_{update}$ = errors during updating steps. Values based on N = 92 participants.

In the case of split attention, analyses display significant results at least for RT_{recall} and RT_{update}. Both coefficient values indicate a small increase in time when attention has to be split up, and descriptive results 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 $Errors_{recall}$, RT_{update} and $Errors_{update}$. Due to the negative coefficient values, results point towards decreased learning performance without the presence of schemata. Descriptive results indeed indicate more 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 RT_{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. *Figure 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.

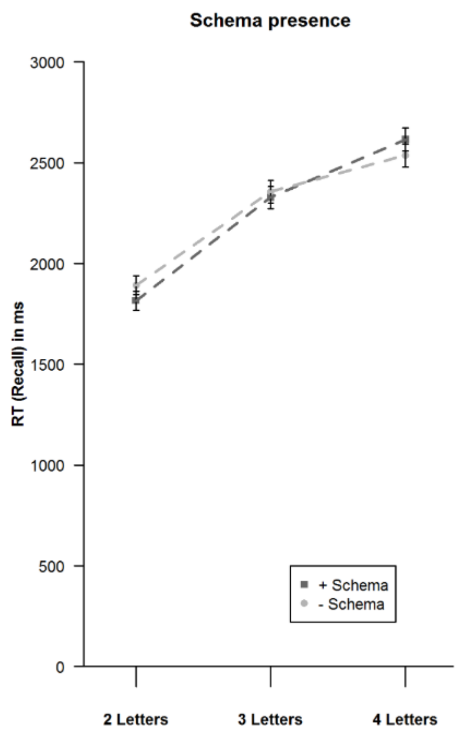

Figure 3. Interaction of complexity and schema presence for RT_{recall}. Dots indicate mean values, error bars indicate standard errors, and dashed lines were inserted to illustrate interactions.

For RT_{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. *Figure 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.

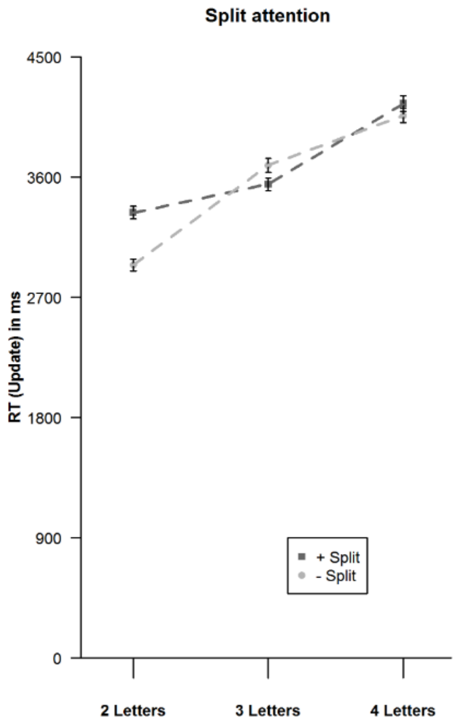

Figure 4. Interaction of complexity and split attention for RT_{update}. Dots indicate mean values, error bars indicate standard errors, and dashed lines were inserted to illustrate interactions.

For $Errors_{update}$, analysis revealed a significant three-way interaction between complexity, split attention, and schema presence. According to *Figure 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.

Table 2

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

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

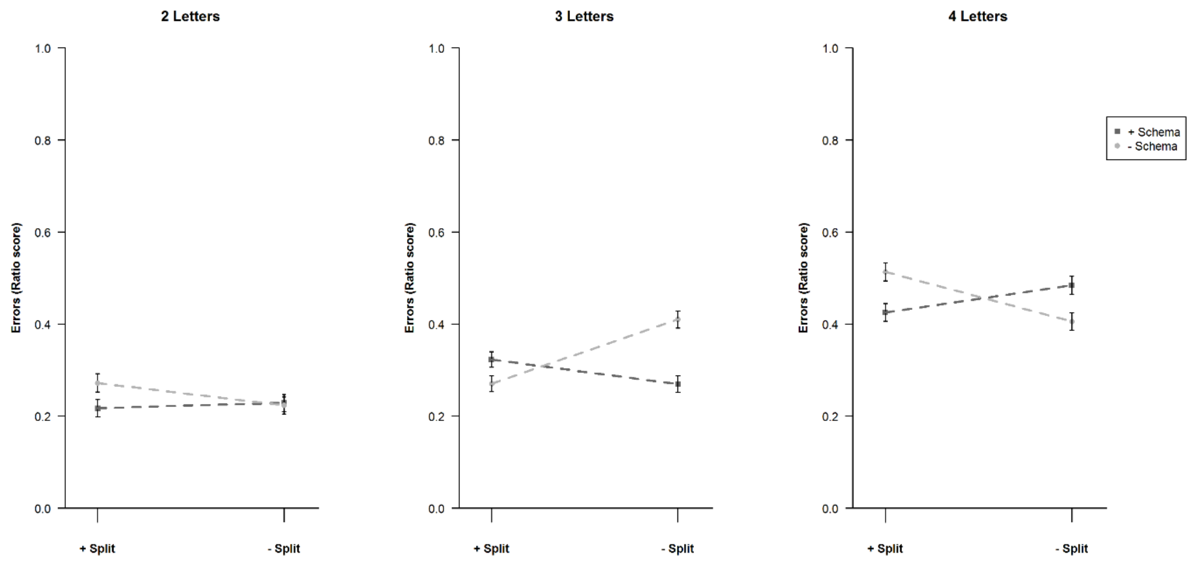

Figure 5. Interaction of complexity, split attention and schema presence for $\text{Errors}_{\text{update}}$. Dots indicate mean values, error bars indicate standard errors, and dashed lines were inserted to illustrate interactions.

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

Analysis 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 $\text{Errors}_{\text{recall}}$, $\text{RT}_{\text{update}}$ and $\text{Errors}_{\text{update}}$, participants achieved smaller values with an increasing level of concentration. Additionally, significant interactions with complexity in the case of $\text{RT}_{\text{recall}}$ and $\text{RT}_{\text{update}}$ showed up, pointing towards stronger differences between conditions in the case of high concentration scores.

For $\text{Errors}_{\text{recall}}$, 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 $\text{RT}_{\text{recall}}$, $\text{Errors}_{\text{recall}}$ and $\text{RT}_{\text{update}}$, since negative coefficients indicate a faster and less erroneous task performance.

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

	RT_{recall} (ms)				**$Errors_{recall}$**				**RT_{update} (ms)**				**$Errors_{update}$**			
	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 x 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 x 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 x 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. RT_{recall} = reaction time during final recall, $Errors_{recall}$ = errors during final recall, RT_{update} = reaction time during updating steps, $Errors_{update}$ = errors during updating steps. Values based on N = 2208 observations of N = 92 participants. [a] df = 90. [b] df = 2105.

Taking a look at the questionnaire 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. Such distinct effects achieved extraordinary strength and significance. 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. 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 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 result 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 the 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).

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 et al. (2010) and Park (2010), indicating interference instead of pure additivity, and correspond well with recent reformulations of the theory (Sweller, 2010; Kalyuga, 2011). 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 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 in 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 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.

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

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.

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

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.

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

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.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed.) (p. 43-71). Cambridge, MA: Cambridge University Press.

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. und 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.

Article 2

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

Original article:

Wirzberger, M., Esmaeili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task, *Acta Psychologica, 179,* 30-41. doi: 10.1016/j.actpsy.2017.07.001

Abstract

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 2 x 5-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.

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

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 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 (Rumelhart, 1980; Bower, Black, & Turner, 1979; Neuschatz, Lampinen, Preston, Hawkins, & Toglia, 2002). 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; Sulin & Dooling, 1974; Brewer & Treyens, 1981). 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 (Rumelhart & Ortony, 1977; Head & Holmes, 1911). 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, 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; Trafton, Altmann, Brock, & Mintz, 2003; Monk, Trafton, & Boehm-Davis, 2008). 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 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 (Yerkes & Dodson, 1908; Ebbinghaus, 1964), raising the need of inspecting plausible nonlinear models of progression (see *Figure 1*).

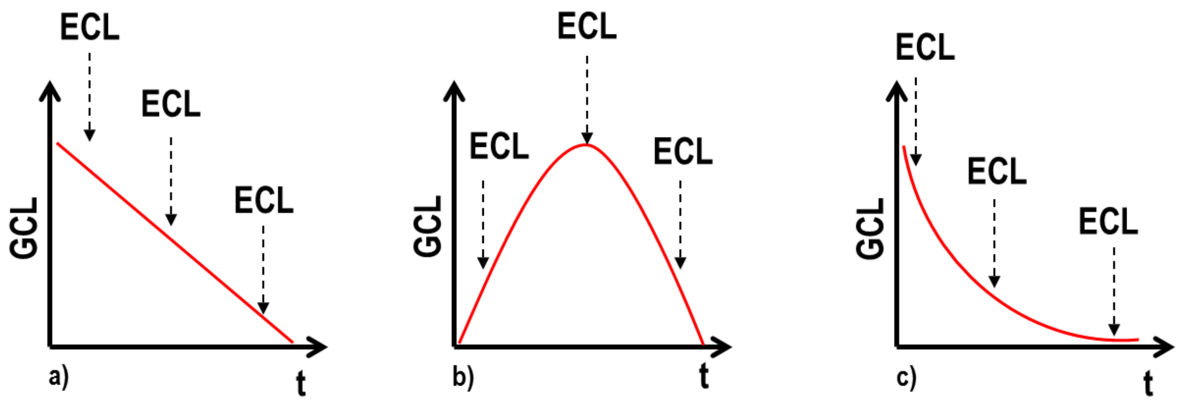

Figure 1. Schematic outline of potential linear and nonlinear temporal models of cognitive load induced during schema acquisition. a) linear progression, b) quadratic progression, c) logarithmic progression.

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 change 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 (Monk et al., 2008; Foroughi, Werner, McKendrick, Cades, & Boehm-Davis, 2016). 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_{age} = 23.25 years, SD_{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 x 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 et al., 2011; Sweller & Chandler, 1994). 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, Redick, Heitz, Broadway, & Engle, 2009; Unsworth, Heitz, Schrock, & Engle, 2005) 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 1,920 x 1,080 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 ($M_{\text{age}} = 33.00$, $SD_{\text{age}} = 15.66$, 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 *Figure 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 & Pylyshyn, 1994; Trick, 2008). However, this span might differ between individuals, depending on practice, age or cognitive skills in enumeration (Trick & Pylyshyn, 1993; Schleifer & Landerl, 2011) 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 13s and 23 s, indicating that further temporal increases in interrupting tasks would not crucially change the arising effect patterns.

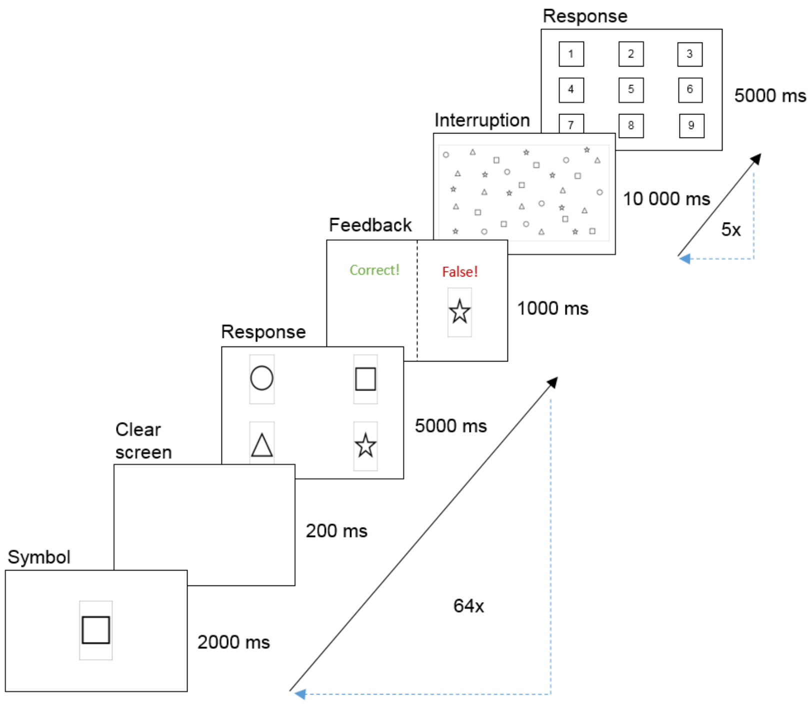

Figure 2. Trial structure within the schema acquisition task. Presentation of second symbol in difficult condition analog with both symbols separated by additional clear screen. Feedback screen according to participant's response (left half: correct choice, right half: false choice; translated version). 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.

2.3.2 *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 x 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 seven-point Likert scale, ranging from not at all (1) to moderately (4) and totally (7).

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, $t(93.92) = -1.01$, $p > .05$, $1-\beta = .78$ (for $d = 0.50$ and $\alpha = .05$), nor by the SSPAN score, $t(105.95) = -0.19$, $p > .05$, $1-\beta = .84$ (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, $F(1, 112) = 34.59$, $p < .001$, $\eta^2_p = .24$. Moreover, previous research found that the longer the duration of an interruption, the greater the time costs. This was also confirmed in this 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. *Figure 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.

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 + $Interaction_{Time(lin)\ x\ Condition}$	Slope $Time_{lin}$ + Participant Intercept	170.85	18450.47	18117.32	196.32	.303
2	$Time_{lin}$ + $Time_{quad}$ + Condition + $Interaction_{Time(lin)\ x\ Condition}$	Slope $Time_{lin}$ + Participant Intercept	173.01	18316.29	17968.03	47.03	.317
3	$Time_{lin}$ + $Time_{log}$ + Condition + $Interaction_{Time(lin)\ x\ Condition}$	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.

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.

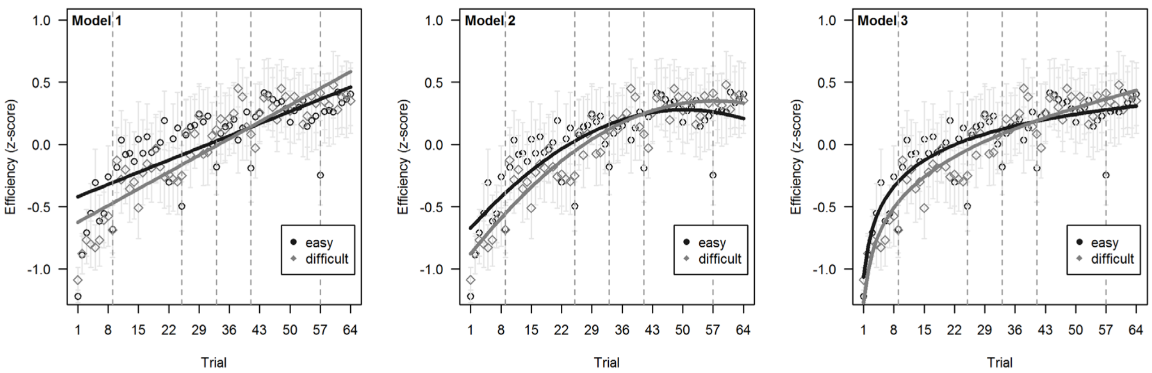

Figure 3. Overall changes in efficiency across trials. Empty dots and rhombs show empirical mean values per trial, lines display predicted values for easy and difficult conditions from chosen models. Dashed vertical lines represent trials following an interruption. Error bars indicate 95% confidence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

Standardized coefficients within the quadratic model indicated a strong significant effect for the linear time predictor (β = .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 (β = -.48, *SE* = 0.04, *t*(7005) = -12.22, p < .001). In addition, the interaction between time and condition showed a small significant effect (β = .05, *SE* = 0.02, *t*(111) = 3.03, p < .05), whereas condition (β = -.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 (β = .32, *SE* = 0.02, *t*(7005) = 14.01, p < .001) and the interaction between time and condition (β = .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 (β = -.02, *SE* = 0.03, *t*(743) = 0.80, p > .05). As in the previous models, condition did not show a significant main effect (β = -.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 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.

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.

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 less than 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 ± 1.29%), the two-way interaction between pre- vs. post-interruption and time over task, $BF_{01} = 1971.564$ (error ± 1.35%) and the three-way interaction between condition, pre- vs. post-interruption and time over task, $BF_{01} = 712.257$ (error ± 1.60%). By contrast, the obtained BF_{01} of 1.739 (error ± 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 $p < .05$). *Figure 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 *Figure 5*.

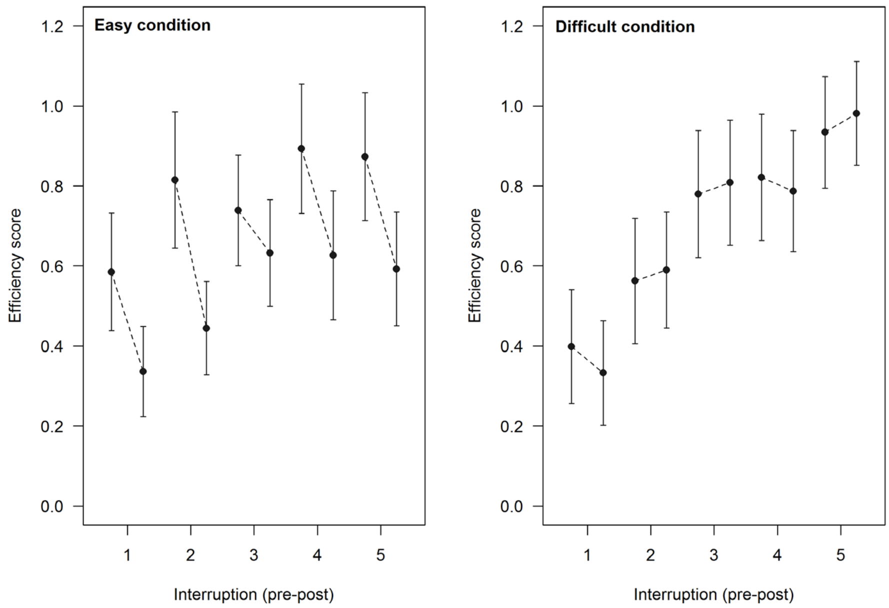

Figure 4. Efficiency in trials immediately before and after an interruption. Dashed lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

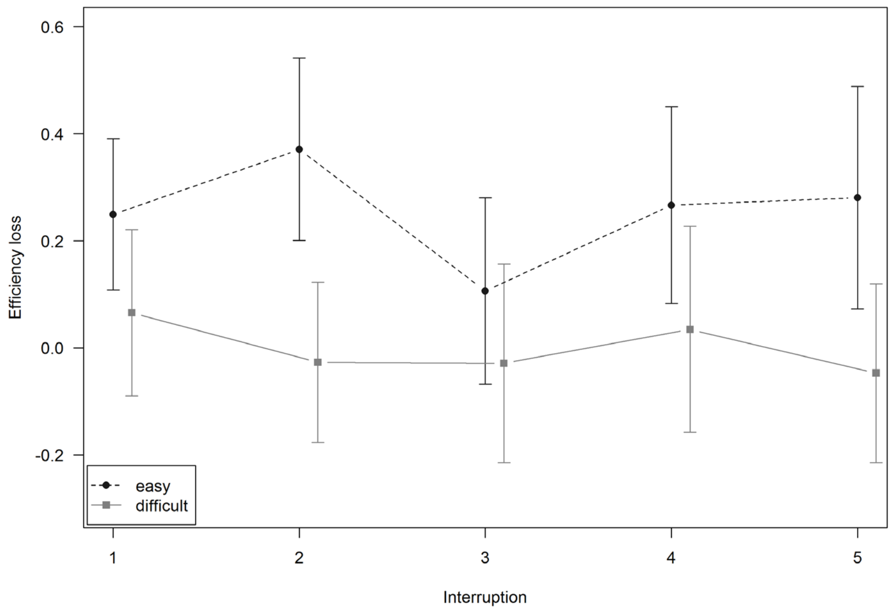

Figure 5. Loss in efficiency due to interruption (pre-post comparison) in easy and difficult conditions. Lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

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 these 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 ± 1.95%), but contradicted the omission of the interaction effect between condition and time of inspection, $BF_{01} = 0.085$ (error ± 0.63%). *Figure 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 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.

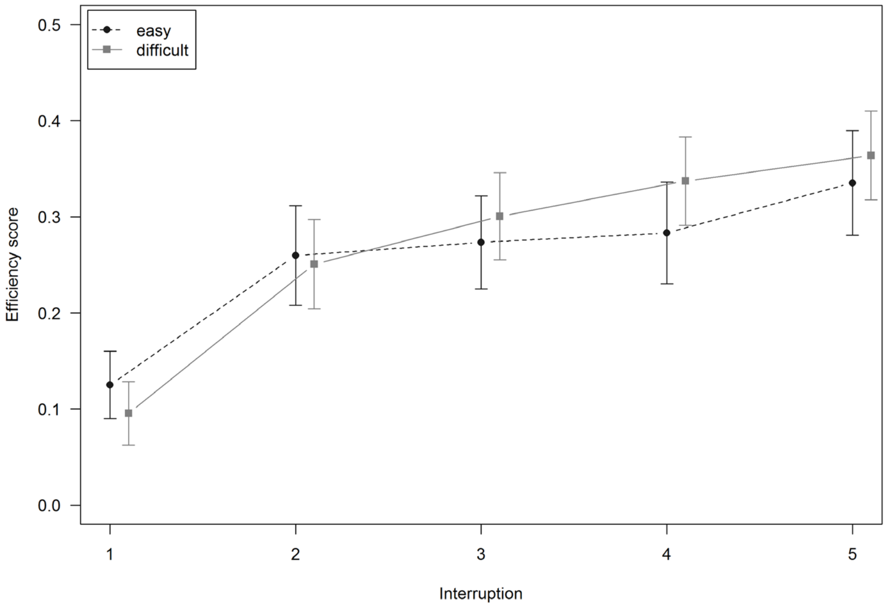

Figure 6. Interruption performance over task. Lines were inserted to facilitate comparisons. Error bars indicate 95% confidence intervals.

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 ($1-\beta = .84$, $\alpha = .05$). 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 ($1-\beta$) 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 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 (Sweller, 2010;

Kalyuga, 2011) 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 & Atkinson, 2003; Renkl, 2014). 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, 1906). 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 warning become less disruptive over time. Comparable to effects of impaired text coherence (McNamara, Kintsch, Songer, & Kintsch, 1996; McNamara, 2001), 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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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

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. doi: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. doi: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. doi: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. doi: 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.* doi: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. doi:10.3758/s13421-014-0461-7

Gagné, R. M., & Dick, W. (1983). Instructional psychology. *Annual Review of Psychology*, *34*, 261–295. doi: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.* doi: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. doi: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. doi:10.1007/BF00309260

Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika, 97*, 773–789. doi: 10.1093/biomet/asq042

Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, *34*, 102–254. doi: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. doi: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. doi: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. doi:10.1080/00223980.1950.9916073

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539. doi: 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. doi: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. doi: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. doi: 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. doi: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. doi: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. doi: 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. doi: 10.1207/s1532690xci1401_1

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. doi: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. doi: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. doi: 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. doi: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. doi: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. doi: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. doi: 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 Anderson R. C., 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. doi:10.3389/fpsyg.2014.00268

Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Developmental science*, *14*, 280-291. doi: 10.1111/j.1467-7687.2010.00976.x

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*, 461–464. doi: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. doi: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. doi:10.1037/h0036846

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1207/s15516709cog1202_4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. doi: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. doi: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. doi: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. doi:10.3758/PP.70.5.743

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. doi:10.1037/0033-295X.101.1.80

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. doi:10.1037//0096-1523.19.2.331

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. doi: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. doi:10.1080/09658210902998047

Valcke, M. (2002). Cognitive load: Updating the theory?. *Learning and Instruction*, *12*, 147-154. doi:10.1016/S0959-4752(01)00022-6

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. doi:10.1016/S0959-4752(01)00019-6

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. doi:10.1016/S0959-4752(01)00020-2

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. doi: 10.1002/acp.1250

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. doi:10.1080/14639220210123806

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4^{th} 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. doi: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. doi:10.1002/cne.920180503

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An Overview. *Educational Psychologist, 25*, 3–17. doi:10.1207/s15326985ep2501_2

Article 3

Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach

Original Article:

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications,* 3:46. doi: 10.1186/s41235-018-0138-z

Abstract

Schema acquisition processes comprise an essential source of cognitive demands in learning situations. To shed light on related mechanisms and influencing factors, this study applied a continuous multi-measure approach for cognitive load assessment. In a dual-task setting, a sample of 123 student participants learned visually presented symbol combinations with one of two levels of complexity while memorizing auditory presented number sequences. Learners' cognitive load during the learning task was addressed by secondary task performance, prosodic speech parameters (pauses, articulation rate), and physiological markers (heart rate, skin conductance response). While results revealed increasing primary and secondary task performance over the trials, decreases in speech and physiological parameters indicated a reduction in the overall level of cognitive load with task progression. In addition, the robustness of the acquired schemata was confirmed by a transfer task that required participants to apply the obtained symbol combinations. Taken together, the observed pattern of evidence supports the idea of a logarithmically decreasing progression of cognitive load with increasing schema acquisition, and further hints on robust and stable transfer performance, even under enhanced transfer demands. Finally, theoretical and practical consequences consider evidence on desirable difficulties in learning as well as the potential of multimodal cognitive load detection in learning applications.

Significance

Interactive learning and training technologies enhance task opportunities in various application domains but may also increase demands on cognitive resources. Arising pitfalls can be avoided by providing systems with knowledge about users' current mental states, which allows regulating system responses in an adaptive and personalized way. For instance, during learning activities, a system could align features such as the amount and speed of the presented content or the degree of instructional support to increase motivation, encourage sustained performance, and foster system acceptance. Related challenges firstly involve measurement issues, i.e., an accurate user state recognition that requires intelligent algorithms for correctly interpreting patterns contained in the acquired signals. Secondly, the system behavior needs to be continuously adjusted to meet users' needs in the sophisticated way possible. The current research provides relevant pre-requisites for both issues by monitoring variations in cognitive demands over the task with a novel combination of sensitive measures related to performance, speech, and physiological reactions. Applying the summarized evidence to real-world applications, dynamic recognition algorithms could be developed for computerized learning devices such as mobile systems that incorporate wearable multimodal sensors.

Keywords: Schema acquisition; Cognitive load assessment; Dual-task setting; Prosodic speech parameters; Physiological measures

1 Background

Looking back in the history of cognitive psychology, there is a long research tradition on cognitive schemata as crucial outcomes of learning processes (Gosh & Gilboa, 2014). Once knowledge has been acquired successfully, it is represented and organized in small bundles of information that are constructed during learning and applied automatically in later process stages. Research in this field has mainly covered structural aspects (Bartlett, 1932; Rumelhart, 1980) and mechanisms of schema acquisition and adjustment (Piaget, 1952), but less concern was devoted to related demands on learners' cognitive resources and their changes during the learning process. The current study addresses this research gap by monitoring the interplay of load inducing factors in a controlled learning scenario with a combination of continuous cognitive load measures.

1.1 Theoretical perspectives in cognitive load research

As we consider cognitive resource investment in instructional situations, the Cognitive Load Theory (CLT; Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998) becomes an indispensable source of explanation. In brief, the theory resides on three core assumptions: Firstly, based on well-established memory models (Anderson, 1983; Atkinson & Shiffrin, 1971; Baddeley, 1992), it postulates limited working memory resources in terms of duration and capacity of information storage. Secondly, long-term memory resources are assumed to lack such boundaries and hold benefits for elaborated learning processes. Thirdly, mental representation of knowledge should occur via schemata, described as organized knowledge structures with stable patterns of relationships between elements (Kalyuga, 2010). A further characteristic of the theoretical framework is the separation of the overall cognitive load construct into the facets of intrinsic, extraneous, and germane cognitive load (Sweller et al., 1998). Whereas intrinsic cognitive load (ICL) results from the number of interrelated elements of information, determining the complexity of the used learning material relative to learners' previous knowledge, extraneous cognitive load (ECL) is associated with the surrounding instructional situation, i.e., ways of content presentation or situational constraints. Germane cognitive load (GCL) arises from relevant processes of schema acquisition and automation. Prior research shows that high levels of ECL hamper learning performance, but only if high amounts of ICL are present at the same time (Sweller et al., 1998). Whereas ECL should be minimized and ICL kept at a manageable level, the instructional focus is put on fostering GCL to achieve optimal learning outcomes.

Theory-related discussions in the more recent past addressed the assumed additive relationship between ICL, ECL, and GCL (De Jong, 2010; Park, 2010; Sonnenfeld & Keebler, 2016) as well as substantial redundancies in the facet of GCL. This facet was introduced in addition to the initial two-component framework mainly on theoretical accounts instead of empirical evidence (Sweller et al., 1998). These issues resulted in efforts on reformulating the postulated theoretical framework. One approach suggests a reduction of cognitive load facets back into a two-component-model, which contrasts productive factors beneficial for learning (ICL) with unproductive factors that impair learning (ECL) and subsumes GCL under the facet of ICL (Kalyuga, 2011; Kalyuga & Singh, 2016; Sweller, 2010). Another approach postulates a process-driven reconceptualization of the three-component-model that quantifies temporal changes in GCL over the task (De Jong, 2010; Sonnenfeld & Keebler, 2016; Wirzberger, Esmaeili Bijarsari, & Rey, 2017). Following this view, ICL, ECL, and GCL reside at different levels of inspection: a structural level in terms of ICL and ECL, which can be determined a priori (Beckmann, 2010; Wirzberger, Beege, Schneider, Nebel, & Rey, 2016), and a processual level in terms of GCL, which undergoes changes throughout the learning task and depends on the achieved level of schema acquisition.

Empirical evidence for the later approach arises from Wirzberger et al. (2017), who applied a basal learning task that a priori varied the amount of interacting information elements (ICL) and induced interruptions at several points over the task (ECL). Statistical analyses compared linear, quadratic, and logarithmic progression models, and results suggested a logarithmic progression of schema-related cognitive load (GCL) over time, influenced by structural features. The resulting inversion of the learning curve (Ebbinghaus, 1964) aligns well with established evidence on cognitive skill acquisition (Anderson, 1983; Kraiger, Ford, & Salas, 1993; Shiffrin & Schneider, 1977). It also provides further evidence that building and organizing schematic structures of knowledge in the initial process stages shed higher demands on cognitive resources, whereas automation and tuning procedures in later process stages need smaller resource supplies. Although the approach already yielded promising results, the study raised the need for a more continuous controlled inspection of the determined mechanisms.

Related questions on the durability and robustness of previously established schemata might be addressable by applying acquired knowledge structures on distinct but related problems. Kalyuga (2010) particularly recommends tasks involving grouping or categorizing to create such transfer demands. Evidence on expertise development shows that novice learners require less complex tasks in early training stages to engage in robust and stable schema acquisition (Van Merriënboer, Kester, & Paas, 2006); thus, novice learners dealing with a complex task right from the beginning should perform worse. In addition, extended transfer requirements would overly demand cognitive resources and further decline performance.

1.2 Approaches to cognitive load assessment

Studying the assumed interplay of cognitive load factors requires their valid assessment. Several efforts have been made, relating to performance, psychophysiology, behavior, and self-report (Sweller et al., 2011; Wickens et al., 2013; Zheng, 2018). From the variety of approaches, a selection will be discussed in the following, which is considered to be relevant for the applied task framework.

Since learners' performance is explicitly addressed and recorded in learning scenarios, the inspection of performance-related parameters offers valuable insights. Such measures usually operate on observable performance indices in dual-task paradigms that use secondary tasks to induce and/or assess cognitive load (O'Donnell & Eggemeier, 1986). If the secondary task mainly serves to induce cognitive load, primary task performance is observed, whereas the inspection of secondary task performance is applied for purposes of assessing cognitive load (Brünken, Plass, & Leutner, 2004; Brünken, Steinbacher, Plass, & Leutner, 2002; Kraiger et al., 1993; Park & Brünken, 2018). A conjoint observation of both aspects provides a more comprehensive view; thus, often both parameters are inspected complementarily. In terms of task-related stimulus and response modalities, evidence on modality compatibility in dual-task performance shows that for spatial codes a combination of visual input and manual output is superior to auditory input and vocal output, whereas results are reversed for verbal codes (Hazeltine, Ruthruff, & Remington, 2006). For this reason, when employing dual-task techniques, compatible input and output modalities for both tasks should be used so that resource interference only occurs due to content-related processing demands. Participants' performance in primary and secondary tasks can further be evaluated in terms of efficiency (Hoffman & Schraw, 2010), with higher levels of efficiency corresponding with high performance and low effort. Efficiency measures usually are calculated from effort indicators like response time, which indicates how cognitively demanding a task was, and performance indicators like correct responses (Sweller et al., 2011).

Bodily functions are often affected involuntarily when people are put under cognitive demands and thus provide reliable online indicators of current levels of cognitive load. Among the variety of psychophysiological techniques, heart rate (HR) and skin conductance response (SCR) have already shown sensitivity to changes in cognitive resource demands in dual-task settings (Mehler, Reimer, & Coughlin, 2012). Both parameters indicate increasing levels of imposed cognitive load by increasing values.

Besides involuntarily occurring bodily responses, participants show voluntarily behavioral reactions as well. Effects of cognitive load on duration-based speech parameters can be classified into the field of prosody, for instance, disfluency, articulation rate, content quality, the number of syllables, and silent pauses as well as filled pauses (Berthold & Jameson, 1999; Müller, Großmann-Hutter, Jameson, Rummer, & Wittig, 2001). Evidence suggests that, with increasing levels of cognitive load, speaking rates (the number of syllables per time) and articulation rates (the number of syllables per time excluding pauses) decrease. More and longer pauses during speech flow, induced by planning processes, also reflect higher levels of cognitive load (Esposito, Stejskal, Smékal, & Bourbakis, 2007; Khawaja, Ruiz, & Fang, 2007, 2008; Müller et al., 2001). So far, the described speech parameters have been applied to determine cognitive load in task settings that demand cognitive resources on a shorter time span, for instance, a reading span task or a Stroop interference task under time pressure (Yap, 2012). A novel perspective arises by applying this approach to capture naturally occurring changes in cognitive resource demands due to schema acquisition processes.

Beneath the introduced objective measures, subjective means of assessment can be applied as well. Self-report rating scales comprise an easily applicable and widely used approach in cognitive load assessment that relies on learners' ability to provide valid retrospective estimations of the experienced level of cognitive load. A recent rating scale that addresses the facets of ICL, ECL and GCL independently was developed by Leppink, Paas, Van der Vleuten, Van Gog and Van Merriënboer (2013). Higher levels of subjectively experienced cognitive resource demands on each facet are indicated by higher ratings on the related items.

In summary, each measure entails certain strengths and weaknesses (Martin, 2018; Sweller et al., 2011; Wickens et al., 2013). Although performance-related parameters provide a continuous measurement and often emerge in any way from the task, secondary tasks potentially interfere with primary tasks and require thoughts regarding the adequate level of complexity as well as the employed stimulus and response modalities. Psychophysiological techniques provide a continuous and reliable measurement, since physiological responses are hardly controllable voluntarily, but require special equipment, substantial expertise, and effort in application and analysis to avoid and control for artifacts and noise. Behavioral parameters also ensure a continuous and reliable way of measurement but require high expertise and effort as well. While subjective ratings via questionnaires are rather easily applicable, they provide no continuous measurement and rely on participants’ retrospective estimations of prior resource demands. For handling the outlined limitations, a combination of different assessment approaches is therefore regarded as the most promising solution to strengthen the informative value of the emerging results (Chen, Zhou, & Yu, 2018; Korbach, Brünken, & Park, 2018).

1.3 Present experiment

To address the outlined research gaps, the current study focused on changes in cognitive load related to the process of schema acquisition, extended by the issue of robustness of the obtained schemata. A basal learning task and a related basal schema application task (Kalyuga, 2010) provided a concise, controllable, and internally valid framework. By combining a set of continuous performance-related, behavioral, and physiological measures, completed by subjective self-reports, the comprehensive capture of underlying cognitive processes was ensured.

1.4 Hypotheses

Based on the introduced theoretical background, a logarithmically decreasing level of cognitive load with increasing schema acquisition was expected, which should be observable in performance, speech, and physiological parameters (*Hypothesis 1a*). Moreover, an increase in subjectively reported cognitive load with a higher level of task complexity was postulated (*Hypothesis 1b*). Regarding retention and transfer performance, with a higher level of task complexity a decrease in retention performance (*Hypothesis 2a*) and transfer performance (*Hypothesis 2b*) was assumed, as well as a decrease in transfer performance with increasing transfer demands (*Hypothesis 2c*).

2 Methods

2.1 Pre-study

A schema application task was designed to address the robustness of participants' schema acquisition. It required participants to categorize symbol combinations with reference to a displayed target symbol as either correct or false, according to previously acquired knowledge. Combinations were composed from the four geometrical symbols square, star, triangle, and circle. Some symbol combinations included non-prototypical symbols to determine if established schemata were stable enough to deal with such kind of transformation. A pre-study should ensure that non-prototypical symbols are still categorized according to their underlying prototype.

2.1.1 Pre-study methods

Seventy-four participants (M_{age} = 35.00 years, SD_{age} = 13.09, range: 19-64, 63.51% female) completed the test, 54.05% already graduated or were currently enrolled as graduate students, 20.27% completed an apprenticeship, 10.81% were undergraduate students, and the remaining 14.87% reported diverse levels of graduation or did not reveal their educational status.

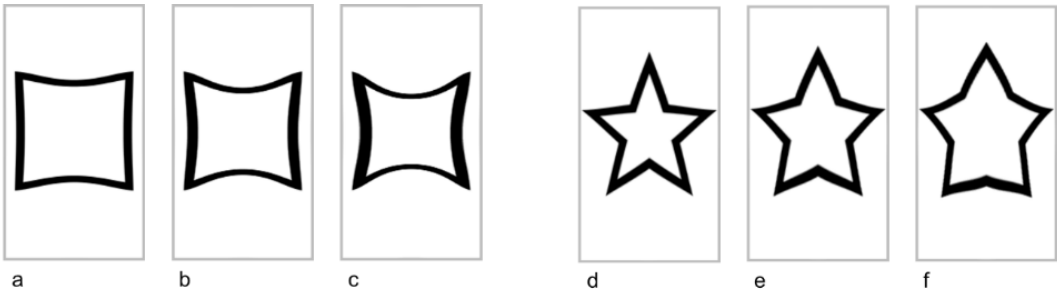

Figure 1. Distorted symbols. a = square 30, b = square 60, c = square 90, d = star -30, e = star -60, f = star -90.

Non-prototypical symbols were generated with Adobe Photoshop and comprised different severities of either barrel (outward, negative deviation from 0) or pincushion (inward, positive deviation from 0) distortion. As displayed in *Figure 1*, levels of distortion ranged from small (±30%) to medium (±60%) to high (±90%).

Participants filled out an online questionnaire including the prototypical and non-prototypical symbols accompanied with a categorization request. After classifying each symbol as either circle, square, star, or triangle, participants had to rate the level of distortion on a seven-point Likert scale with verbal anchoring at the extreme points to determine the perceived severity of deviation from the underlying prototype. As the prototypes were included in the presented symbol set as well, the scale started from 0 (“not at all”) and reached to 6 (“very strong”) to provide participants the opportunity of a valid and reliable rating on a sufficient level of complexity (Lozano, García-Cueto, & Muñiz, 2008).

2.1.2 Pre-study results

Descriptive analyses of categorization frequencies and distortion ratings are reported in *Table 1*. They indicate rather stable and homogeneous classifications of non-prototypical symbols according to their underlying prototype, with at least 94.7% correct categorizations, even on the broader distribution of age and educational backgrounds. Moreover, even on a descriptive level, distortion ratings outline an increasing amount of perceived deviation with increasing severity of distortion.

Repeated-measures analyses of variance (ANOVAs) on the distortion ratings were computed separately for each symbol category, with severity of distortion as sevenfold within-subjects factor. Mauchly's test indicated a violation of sphericity for all symbol categories; therefore, the degrees of freedom were corrected using Greenhouse-Geisser estimates for circle, $\chi^2(20) = 69.184$, $p < .001$, $\varepsilon = 0.751$, square, $\chi^2(20) = 61.139$, $p < .001$, $\varepsilon = 0.778$, star, $\chi^2(20) = 156.934$, $p < .001$, $\varepsilon = 0.608$, and triangle, $\chi^2(20) = 49.509$, $p < .001$, $\varepsilon = 0.837$. The results indicated strong and highly significant main effects of distortion severity for circle, $F(4.504, 388.761) = 109.263$, $p < .001$, $\eta_p^2 = .599$, square, $F(4.668, 340.729) = 151.472$, $p < .001$, $\eta_p^2 = .675$, star, $F(3.649, 266.373) = 87.503$, $p < .001$, $\eta_p^2 = .545$ and triangle, $F(5.020, 366.476) = 175.157$, $p < .001$, $\eta_p^2 = .706$.

Table 1

Categorization and distortion rating of prototypical and distorted symbols

Dist.	Circle			Square			Star			Triangle		
	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*
0	100	1.09	0.38	100	1.12	0.72	97.3	1.03	0.16	100	1.03	0.16
-30	100	2.45	1.04	100	2.82	1.03	100	1.49	0.82	100	2.28	0.77
30	98.7	2.20	0.98	98.7	2.76	1.07	100	1.27	0.45	100	2.81	1.18
-60	100	3.27	1.31	98.7	3.77	1.17	100	2.14	1.24	98.7	3.14	1.16
60	100	3.18	1.33	100	3.72	1.28	100	2.08	1.16	94.7	4.03	1.24
-90	100	3.69	1.33	100	4.65	1.21	100	2.32	1.25	100	3.46	1.28
90	98.7	3.70	1.32	97.3	4.53	1.40	98.7	3.61	1.50	98.7	4.59	1.32

Note. Dist. = severity of out- or inward distortion; % = percentage of correct categorizations; *M* = mean of distortion rating; *SD* = standard deviation of distortion rating.

Post-hoc pairwise comparisons with Bonferroni-Holm correction (Maxwell, 1980) revealed significant differences between distorted symbols and prototypes for all levels of distortion severity in each symbol category ($p < .001$). When comparing out- and inward distortion within and across distortion severities, diverse patterns showed up depending on the symbol category. For both circle and square, no significant differences resulted between 30 vs. -30, 60 vs. -60 and 90 vs. -90, whereas all other comparisons achieved significance with at least $p < .05$. In the star category, no significant differences resulted between 60 vs. -60, 60 vs. 90, and -60 vs. 90, but significant differences ($p < .001$) occurred for all other comparisons. All pairwise comparisons achieved significance in the triangle category with at least $p < .05$.

Based on both categorization frequencies and distortion ratings, non-prototypical symbols with superior psychometric properties were chosen. In addition to considering correct classifications and significant pairwise comparisons, a balanced representation of both barrel and pincushion distortion across symbols was sought. The resulting pattern comprised -30, -60, -90 for circle, -30, 60, -90 for square, 30, 60, -90 for star, and 30, -60, -90 for triangle.

2.2 Main study

2.2.1 Participants

A total of 123 undergraduate and graduate students (M_{age} = 22.67 years, SD_{age} = 3.55, range: 18-34, 76.42% female) participated in the main study. They were enrolled in Communication Science (41.32%), Psychology (30.58 %), STEM fields (11.57%), Humanities (9.09%) or Education (5.79%) and received either a financial allowance of 5 € (49.59%) or course credits (50.41%) as compensation. Experimental conditions did not differ regarding age, $t(119.05) = 0.62$, $p = .539$, $d = 0.111$, gender, $\chi^2(1) < 0.01$, $p = .960$, distribution of study courses, $\chi^2(6) = 4.42$, $p = .620$, or compensation choice, $\chi^2(1) = 0.40$, $p = .527$.

2.2.2 Design

The chosen learning task required participants to detect, remember and retrieve four easy or difficult combinations of arbitrary geometric symbols while simultaneously memorizing five-digit number sequences as a secondary task. The resulting experimental design included task complexity as independent between-subjects factor that varied due to the arrangement of the symbol combinations (three vs. four symbols, symbol order). As dependent variables, learners' performance in both primary and secondary task, spoken responses on the secondary task, and physiological reactions were recorded continuously during the learning task. The standardized cognitive load questionnaire by Leppink et al. (2013) provided a summative evaluation of the inspected cognitive load facets. An open question on schema recall after the learning task enabled insights into the quality of schema acquisition over the task. Participants' working memory capacity was derived from a translated version of the automated operation span task (OSPAN; Unsworth, Heitz, Schrock, & Engle, 2005) and used as a control variable.

With reference to the CLT, task complexity reflected the ICL component and was addressed according to the concept of element interactivity (Sweller, 2010). Unlike the work of Wirzberger et al. (2017), the scope of symbol combinations was increased by one element to avoid effects of boredom due to insufficiently low levels of task demands. The embedded secondary task characterized the ECL component and aligns with the conceptualization of ECL as situational constraints (Wickens et al., 2013). Furthermore, the combined inspection of the continuous cognitive load measures hinted at the underlying cognitive resource investment pattern and represented the GCL component (Sweller et al., 2011). Whereas primary task efficiency resources invested in schema acquisition relative to the achieved outcome (Hoffman & Schraw, 2010; Paas & Van Gog, 2006), secondary task efficiency hinted on the actual load imposed by the primary task. Based on this view, available resources arise only due to already established and usable schemata (Kraiger et al., 1993).

An additional task on applying the obtained schemata by solving a categorization task addressed the robustness of the schema acquisition process. It required participants to evaluate the correct match of displayed input and response parts of the previously acquired symbol combinations, which included distorted symbols in parts of the trials. The underlying 2 x 5 factorial mixed design included a between-subjects factor task complexity (easy vs. difficult) and a within-subjects factor level of distortion (0 vs. 30 vs. 60 vs. 90) as independent variables. Reaction time and a corrected error score were recorded as dependent variables.

2.2.3 Materials

2.2.3.1 Learning task

Computer-based tasks were realized with OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and provided on a standard desktop computer with a 24" monitor. Trials within the learning task were inspired by the procedure of the OSPAN (Unsworth et al., 2005) and comparable automated complex span tasks, which include a distractor task and a target task in each trial in alternating sequence. In the current study, within each trial the number task (secondary task) was presented first as a distractor and followed by the symbol task (primary task) as a target. Unlike the complex span task procedure, which involved new items to memorize in each trial for both the target and distractor task, the learning content of the primary task persisted during the entire task.

Each of the 64 trials started with the auditory presentation of a unique randomly chosen five-digit number for 5000 ms (see *Figure 2*). The amount of five digits, as well as the used time spans, was determined within a short internal pretest with N = 7 participants (M_{age} = 28.71, SD_{age} = 2.43, range = 26-32, 4 male) and chosen to avoid distraction effects on the primary task by an overly complex secondary task. Indicated by a black speaker symbol on the screen, participants had to listen carefully and memorize the numbers in correct sequence. Afterward, a randomly chosen input part of one out of four abstract symbol combinations was presented for 2 s. This input part comprised two symbols in the easy and three symbols in the difficult conditions. Participants had to remember the shown symbols in correct order and complete the sequence by choosing a symbol as response on the next screen by mouse click. In this vein, a total of three (easy condition) or four (difficult condition) symbols formed a complete combination.

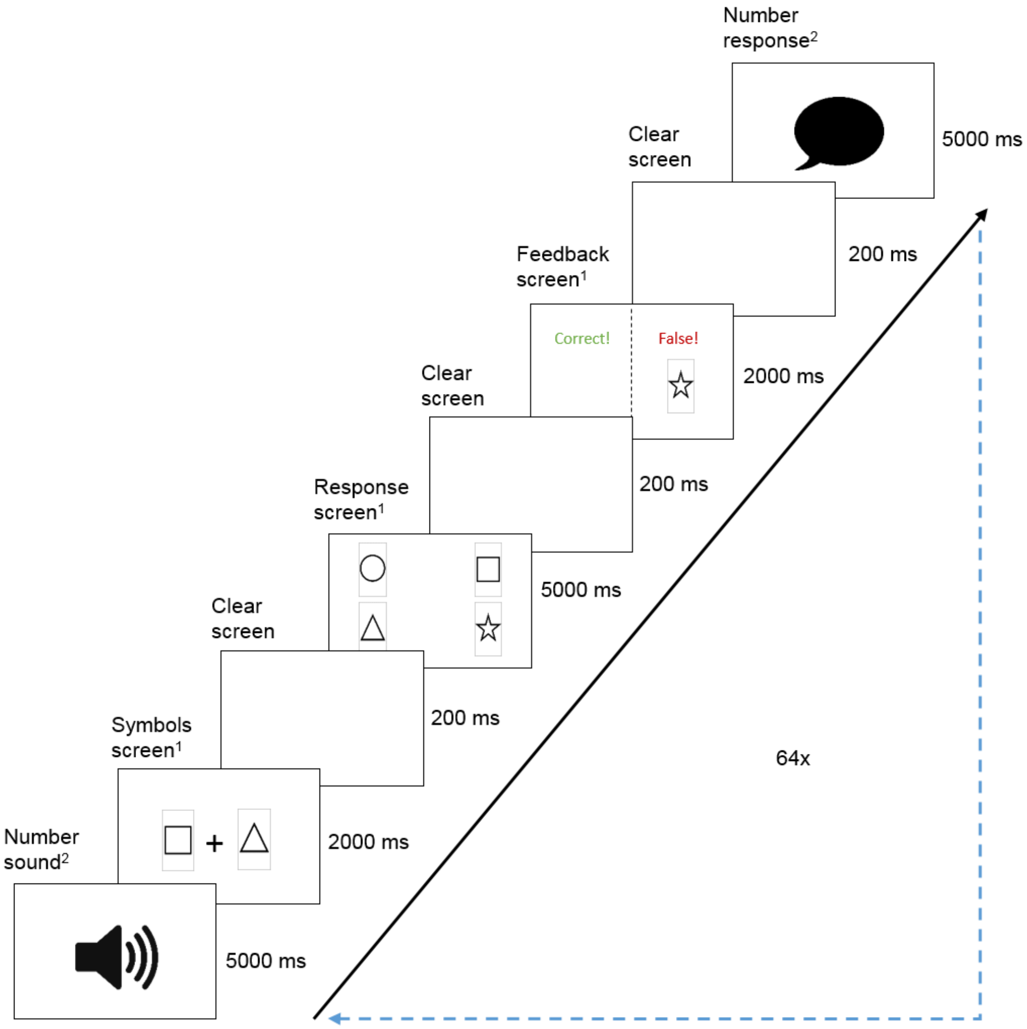

Figure 2. Schematic outline of the trial procedure in the learning task. Superscript indices indicate affiliation with the primary[1] or secondary[2] task.

As shown in *Figure 2*, the response screen simultaneously presented the four possible response symbols in a randomly arranged 2 x 2 grid for 5 s. This was followed by a feedback screen lasting 2 s that also included the correct choice for false responses to foster correct schema acquisition. Finally, indicated by a black speech bubble symbol, participants had to recall the memorized digit sequence from the trial outset verbally within 5 s.

2.2.3.2 Schema application task

During each of the 60 trials in the schema application task, participants evaluated if the input part of a presented symbol combination matched or mismatched the response part. As depicted in *Figure 3*, the response part of the symbol combination was shown in the upper part and the potential input part in the lower part of the screen. Response parts were always represented in prototypical symbols, whereas half of the input parts included distorted symbols from the pre-study. The pool of presented stimuli comprised correctly matched input and target parts, existing input parts with mismatched target parts and non-existing input parts with mismatched target parts. Within 5 s, participants had to classify the presented combination as false or correct by pressing either the left (false) or right (correct) mouse key. Contrary to the learning task, participants received no further feedback on their response.

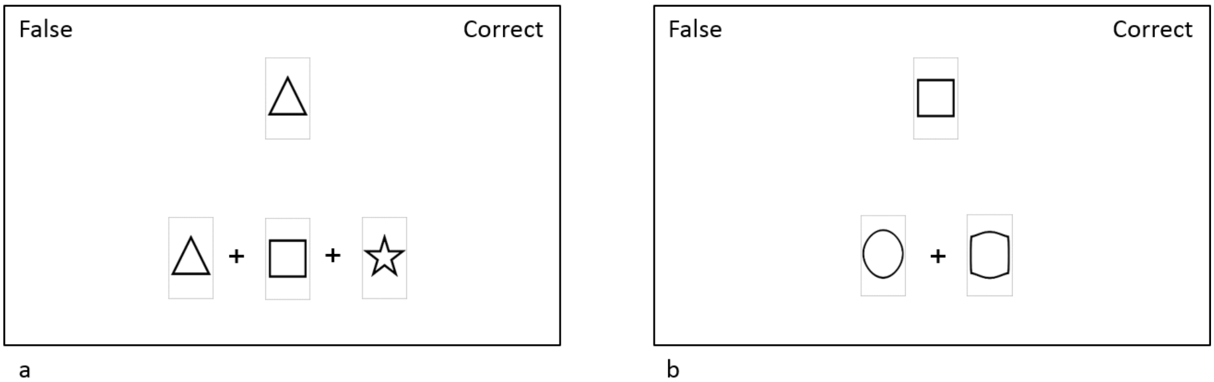

Figure 3. Stimuli in the schema application task: (a) difficult condition with prototypical symbols and (b) easy condition with distorted symbols.

2.2.3.3 Questionnaires on retention and cognitive load

For assessing retention performance, participants received a computer-based form with a grid of three (easy condition) or four (difficult condition) empty boxes per combination in four rows. They had to recall the memorized symbol combinations by dragging symbols from an infinite pool for each possible symbol (i.e., circle, square, triangle, or star) via mouse click and dropping them into the existing grid boxes to form combinations. In addition, they could indicate corrections on the provided combinations in a separate comment space below. Participants' subjectively perceived level of cognitive load was addressed on 11-point Likert scales via the 10-item questionnaire by Leppink et al. (2013) that differs between the three subscales of ICL, ECL, and GCL.

2.2.4 Procedure

Data were obtained in individual testing sessions of about 60 min in a separate laboratory, equipped with a standard desktop computer for the participant and an experimenter netbook to record the physiological data. At the outset of each testing session, participants were welcomed and signed an informed consent. This form outlined the purpose and procedure of the study and ensured that participants' treatment aligned with approved ethical standards and their privacy was respected. Afterward, the OSPAN had to be completed, which usually took about 15 min.

Before starting the learning task, the experimenter had to fulfill several preparatory duties, requiring about 5 min. First, the electrodes for the physiological measures were attached to the left hand and a close-talk microphone for the speech recording was placed at the upper part of the participant's sternum. Following an initial adjustment of the speech recording volume to individual voice characteristics, the learning task was completed within an average duration of 20 min. After the electrodes were removed, participants worked on the questionnaires on retention, cognitive load, and demographics, which were provided online and were completed in about 10 min by most participants. Finally, they completed the schema application task (about 10 min) and were debriefed and approved.

2.2.5 Scoring

Primary task efficiency was computed, following the likelihood model (Hoffman & Schraw, 2010), as quotient of correct responses (performance) and reaction times (effort) within each trial. Since reaction times were retrieved in milliseconds, scores were multiplied by 1000 to obtain the proportion of correct responses per second. The resulting values provide a hint on the investment of available mental resources over the task: if learners start to perform faster and less erroneous on the task, they must invest less mental capacities.

For *secondary task efficiency*, also based on the likelihood model approach (Hoffman & Schraw, 2010), performance was computed by comparing spoken words to correct words from the reference and subtracting the amount of substituted, deleted, and inserted words (Lee, 1988) relative to the total number of words in the reference (word accuracy). The participant's effort was reflected in the time starting from the presentation of the visual stimulus to the end of the last uttered digit (verbal response duration), which covered the entire answer process.

Speech parameters were extracted on phoneme level using an automatic speech recognition system (Herms, 2016). The resulting transcripts included spoken units and the corresponding time codes in milliseconds and were used to derive the articulation rate, the number, and the mean duration of silent pauses. The *articulation rate* represented the total number of phonemes divided by the utterance duration excluding the total duration of silent pauses, the *number of pauses* reflected the total number of silent pauses in an utterance, and the *mean pause duration* was calculated from the total duration divided by the number of silent pauses.

Physiological data were recorded at a frequency of 128 Hz with a NeXus-10 Mark II from sensors attached to the volar surface of the distal phalanges of the left hand. While the heart rate (HR) signal was obtained at the trigger finger, the skin conductance response (SCR) was recorded at the middle and ring finger. Data preparation involved the calculation of an individual baseline for each participant from values located in the preparation phase of about 5 min before starting the learning task. Recorded *SCR* and *HR* values were normalized to the individual baseline value and aggregated on mean values for events within each trial. Each event represented a subtask within the learning trial and was related to a screen change, i.e., the auditory presentation of a number sequence, the visual presentation of the symbol combination input, the response part of the presented symbol combination, a feedback on the given response, and finally the verbal recall of the memorized number sequence.

For obtaining *retention performance*, sum scores were calculated for all memorized symbol combinations and all correctly memorized symbol combinations, resulting in values ranging from 0 (neither combination memorized correctly) to 4 (all combinations memorized correctly). *Transfer performance* was obtained from the schema application task in terms of reaction times on the provided classifications as well as correct responses. The latter were adjusted for inherited errors from the retention task.

Cognitive load scores were derived from sum scores for each cognitive load facet, resulting in a maximum of 30 points for ICL and ECL, and a maximum of 40 points for GCL. Subscales achieved satisfying internal consistencies of $\alpha = .831$ for ICL, $\alpha = .708$ for ECL and $\alpha = .876$ for GCL. In line with Conway et al. (2005), who reported a clear advantage of partial credit scoring procedures over all-or-nothing scoring procedures, the partial load score for the *OSPAN* was computed by awarding one point for each correctly recalled letter. Across the three test blocks, the task achieved an appropriate internal consistency of Cronbach's $\alpha = .805$.

3 Results

3.1 Cognitive load progression

Some datasets had to be excluded from data analysis due to missing values or the failure to meet the 85% accuracy criterion in the OSPAN[1] task, a lack in language proficiency, or an observable violation of the instruction. Conditional growth curve models were computed to inspect progressions in primary and secondary task efficiency from a temporal perspective. Values for all relevant variables were *z*-standardized to obtain standardized β coefficients. Models were fit with restricted maximum-likelihood estimation and included time, condition, the OSPAN score, and the interaction between time and condition predictors as fixed effects aligned to the experimental task design. To take into account individual differences between participants in reaction to experimental variations, a time slope as well as subject-specific

[1] This exclusion involved seven participants, which corresponds to about 6% of the sample. For comparison, Unsworth et al. (2005) report exclusion rates of 15%.

intercepts were considered as random effects and assumed to be correlated. In addition, time was computed as logarithmic[2] variable and included as fixed effect. For evaluating model fit, the root mean squared error (RMSE) was obtained from a leave-one-out-cross-validation approach and the conditional pseudo R^2 for generalized linear mixed models was calculated.

3.1.1 Primary task efficiency

After exclusions, the analysis of primary task efficiency operated on $n = 103$ datasets. As indicated by *Figure 4*, significant contributions resulted for the linear time predictor, $\beta = .166$, standard error (SE) = 0.030, $t(683) = 5.609$, $p < .001$, the logarithmic time predictor, $\beta = 0.118$, $SE = 0.026$, $t(6385) = 4.573$, $p < .001$, and condition, $\beta = -.070$, $SE = 0.031$, $t(100) = -2.220$, $p = .029$. No significant interaction between time and condition could be observed, $\beta = -.031$, $SE = 0.018$, $t(101) = -1.726$, $p = .087$. The model achieved an acceptable fit with RMSE = .936 and $R^2 = .197$ and supports *hypothesis 1a*.

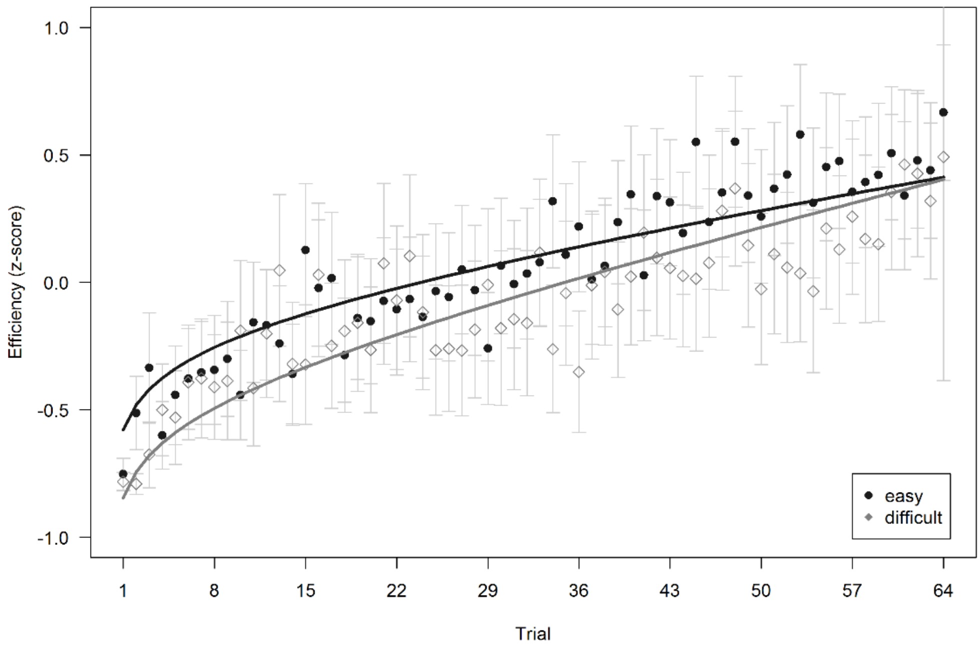

Figure 4. Changes in performance in primary task efficiency over the task. Filled dots and empty rhombs show empirical mean values per trial, lines indicate predicted mean values. Error bars indicate 95% confidence intervals from empirical observations.

3.1.2 Secondary task efficiency

When analyzing secondary task efficiency, one additional dataset had to be excluded from the sample due to high noise in the recorded speech signal, resulting in a subsample of $n = 102$ participants. Corresponding to the prevention of distraction effects on the primary task from an

[2] In line with Wirzberger et al. (2017), alternative linear and quadratic progression models were tested as well, but not reported due to the confirmed advantage of the logarithmic model and the lack of benefit for the focus of this study.

overly complex secondary task, participants achieved a predominantly high word accuracy (*M* = 0.95, *SD* = 0.15). As visually supported by *Figure 5*, results revealed an increasing secondary task performance efficiency over time as well. In more detail, standardized coefficients showed a significant logarithmic time predictor, $\beta = .160$, $SE = 0.022$, $t(6323) = 7.392$, $p < .001$, whereas no significance could be obtained for the linear time predictor, $\beta = .029$, $SE = 0.028$, $t(361) = 1.022$, $p = .307$, neither the effect of condition, $\beta = -.044$, $SE = 0.062$, $t(99) = -0.711$, $p = .479$, nor the interaction between time and condition, $\beta = -.006$, $SE = 0.021$, $t(100) = -0.271$, $p = .787$. The overall model achieved a considerable fit with RMSE = .912 and $R^2 = .445$ and supports *hypothesis 1a*.

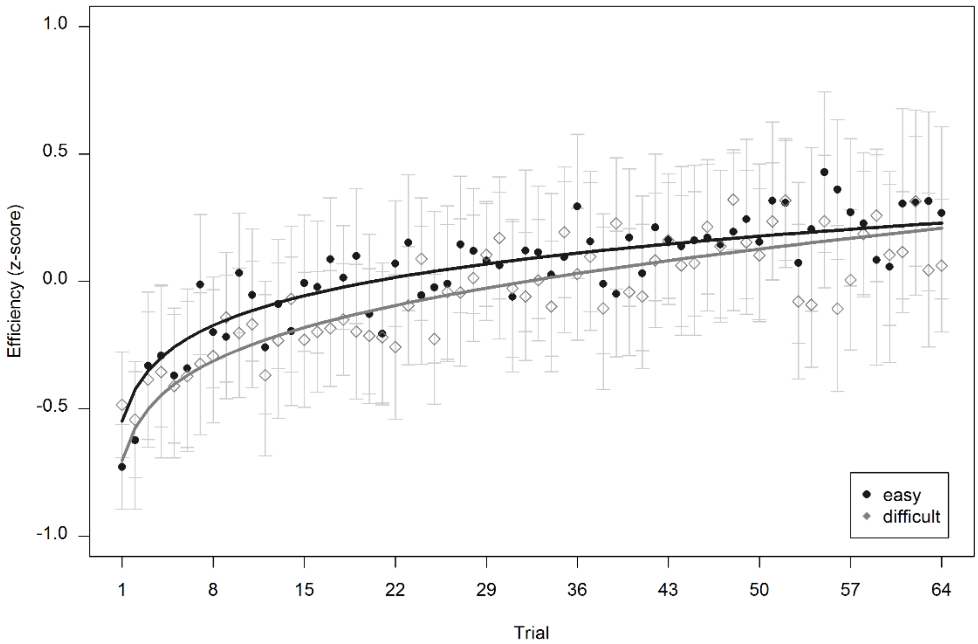

Figure 5. Changes in performance in secondary task efficiency over the task. Filled dots and empty rhombs show empirical mean values per trial, lines indicate predicted mean values. Error bars indicate 95% confidence intervals from empirical observations.

3.1.3 Speech-related parameters

Corresponding to secondary task performance, results are based on $n = 102$ participants. Time series regressions with linear and logarithmic trend predictors, separated by conditions as depicted in *Figure 6*, support decreases in cognitive load over time in terms of mean pause duration, $R^2_{easy} = .063$ and $R^2_{difficult} = .216$. In terms of number of pauses and articulation rate, the models achieved $R^2_{easy} = .131$ and $R^2_{difficult} = .311$ for number of pauses and $R^2_{easy} = .165$ and $R^2_{difficult} = .268$ for articulation rate. Although amounts of explained variance differ between measures and conditions, the overall trend supports *hypothesis 1a*.

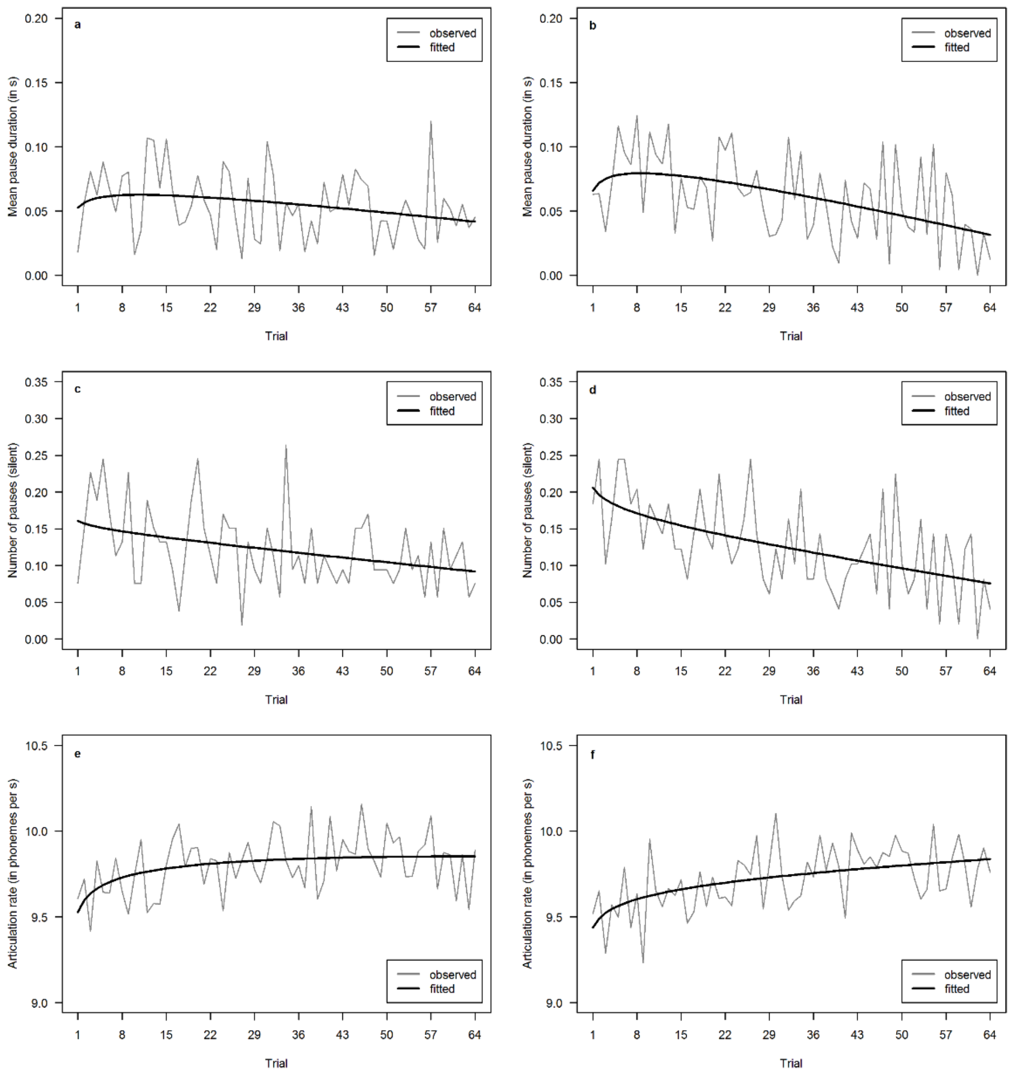

Figure 6. Observed and fitted logarithmic progressions of speech-related parameters computed per trial in easy (a, c, e) and difficult (b, d, f) conditions. Graphs in the first row refer to mean pause duration (total duration of silent pauses in seconds divided by the number of silent pauses). Graphs in the second row refer to the number of pauses (only silent pauses). Graphs in the third row refer to articulation rate (number of phonemes per second excluding pauses).

3.1.4 Physiological parameters

Due to technical errors in the recorded data, two additional datasets had to be excluded from the analysis, resulting in a subsample of n = 101 participants. Assuming additive time series with trend and seasonal components, analyses indicate a decreasing trend of HR and SCR over the task and show a repetitive seasonal pattern across “subtasks” within each trial. Logarithmic time series regression models, including linear and nonlinear trend predictors as well as seasonal predictors, achieved R^2_{easy} = .847 and $R^2_{difficult}$ = .672 for SCR, whereas for HR an R^2_{easy} = .590 and $R^2_{difficult}$ = .643 could be obtained. Inspecting the seasonal component in more detail, following an initial increase while the sequence of numbers was presented auditorily, a decreasing progression over the symbol presentation, symbol response and symbol feedback could be observed. However, in the final step of verbally recalling the memorized numbers, physiological response parameters increased again. *Figure 7* displays the respective observed and fitted logarithmic progression curves, which strongly support *hypothesis 1a*.

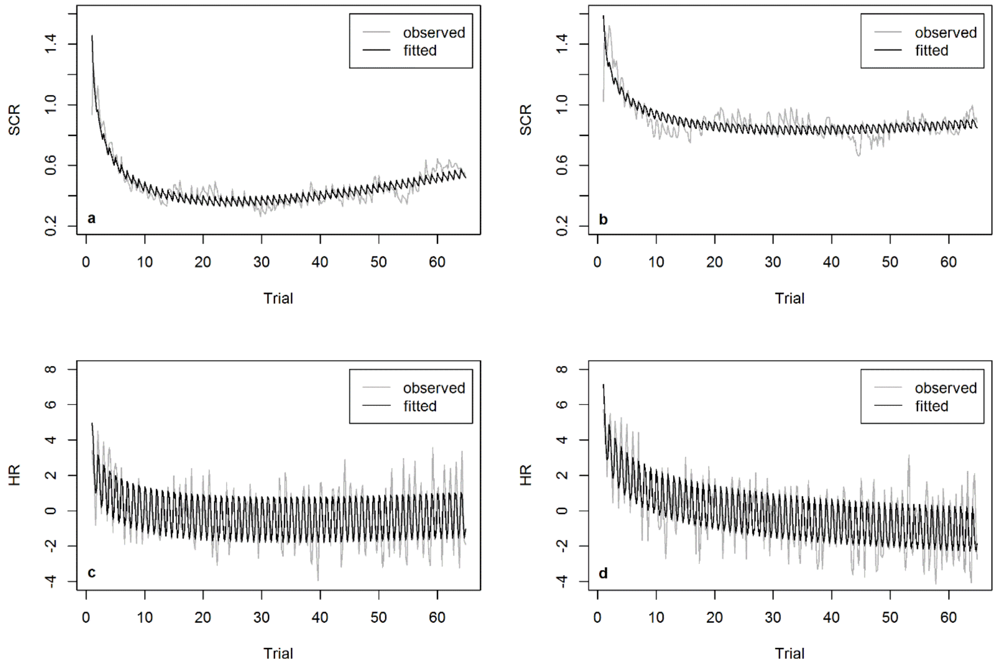

Figure 7. Observed and fitted logarithmic progressions of SCR and HR over time in easy (a, c) and difficult (b, d) conditions including both seasonal and trend components.

3.2 Retention and transfer performance

3.2.1 Retention performance

Results for both retention and transfer performance are based on n = 103 participants. In terms of retention performance, in line with *hypothesis 2a*, significantly fewer correctly recalled symbol combinations could be observed in the difficult conditions, $t(96.53) = 4.72$, $p < .001$, $d = -0.92$. No significant differences between conditions regarding the amount of totally recalled symbol combinations resulted, $t(100.99) = -0.08$, $p = .937$, $d = 0.02$. A power[3] level of $1-\beta = .71$ was achieved for $\alpha = .05$ and $d = 0.5$.

3.2.2 Transfer performance

As already indicated from the descriptive values in *Table 2*, results revealed significantly faster responses in the difficult condition, $F(1,101) = 5.59$, $p = .020$, $\eta_p^2 = .05$, reversing the pattern postulated in *hypothesis 2b*. The significant increase in reaction time with increasing distortion,

[3] Power analyses refer to theoretically assumed population effect sizes to provide a broader informative value for the reasoning about non-significant results (O'Keefe, 2007).

$F(3,303) = 15.13$, $p < .001$, $\eta_p^2 = .13$, partially supports *hypothesis 2c*. A significant interaction between both factors did not show up, $F(3,303) = 0.60$, $p = .614$, $\eta_p^2 = .01$, 1-β = 1.0 for α = .05 and $f = .25$. Post hoc pairwise comparisons using Tukey's honest significant difference (HSD; Maxwell, 1980) indicate significant differences in reaction time between the prototypical level and all levels of distortion (all *p*'s < .05) as well as the distortion levels 30 and 90 ($p < .001$). After correction for inherited errors, no significant differences result in correct responses between conditions, $F(1,101) = 0.99$, $p = .323$, $\eta_p^2 = .01$, 1-β = .941, and levels of distortion, $F(3,303) = 1.42$, $p = .237$, $\eta_p^2 = .01$, 1-β = 1.0. Likewise, no significant interaction between both factors could be detected, $F(3,303) = 0.91$, $p = .436$, $\eta_p^2 = .01$, 1-β = 1.0. All reported power levels relate to α = .05 and $f = .25$. In general, with around 70% in each level of distortion, a high frequency of correct responses was observable (see *Table 2*), hinting on a rather stable and robust application of previously acquired schemata.

Table 2

Descriptive values of correct responses and reaction time for levels of distortion in conditions

Dist.	**RT_{easy}**		**$RT_{difficult}$**		**CR_{easy}**		**$CR_{difficult}$**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0	2571.14	1079.33	2261.58	1010.11	0.75	0.44	0.71	0.46
30	2692.68	1168.68	2381.37	1116.81	0.70	0.46	0.70	0.46
60	2670.05	1208.94	2485.99	1091.26	0.72	0.45	0.69	0.46
90	2776.47	1178.77	2519.48	1097.61	0.72	0.45	0.68	0.47

Note. Dist. = severity of distortion in %; RT = reaction time; CR = correct responses (min = 0, max = 1); easy = easy condition; difficult = difficult condition.

3.3 Subjective cognitive load ratings

Contrary to *hypothesis 1b*, no differences between easy and difficult conditions showed up for ICL, $t(90.76) = -1.13$, $p = .225$, $d = 0.306$, and ECL, $t(97.57) = -1.22$, $p = .226$, $d = 0.242$, although descriptive values pointed towards lower scores for the easy condition. By contrast, significantly higher ratings in the easy condition resulted for GCL, $t(98.77) = 2.87$, $p = .005$, $d = -0.568$, which directly reverses the pattern postulated in *hypothesis 1b*. According to Cohen (1988), the effect amounts to a medium size. Analyses achieved a power of 1-β = .709 for α = .05 and $d = 0.5$.

4 Discussion

This study provided insight into the progression and interaction of the outlined facets of cognitive load by applying a multi-method approach to cognitive load assessment. In line with the postulated hypothesis, performance, speech, and physiological parameters indicated a logarithmically decreasing level of cognitive load over the task, hinting at increasing progress in schema acquisition. The subjective ratings did not support the initial assumptions, but support arises for the decrease in retention performance with higher levels of task complexity. For transfer performance, the stated hypothesis on increased transfer demands was partly confirmed in terms of reaction times, while evidence indicated a reversed pattern for task complexity in this case.

In addition, physiological measures revealed a repetitive seasonal pattern across the subtask routines for SCR and HR: during the presentation of the auditory stimulus at the outset of each trial, an increase was observable. It was followed by a decrease related to the visual presentation of the symbol combination, the motor response selection, and the visual feedback screen. As soon as the verbal response on the initially presented digit sequence was requested, again increasing levels in both physiological measures resulted. This evidence suggests higher levels of cognitive load imposed by the auditory-verbal compared to the visual-motor stimulus-response combination due to the observable initial and final increase in the emerging signal. Support for this assumption arises from Posner, Nissen, and Klein (1976), who state the dominance and increased familiarity of the visual modality across from other modalities. Another potential clarification suggests increased perceptual load (Lavie, 2010) in the auditory-verbal secondary task, imposed by the additional visual stimulus from the speaker and speech bubble symbols shown on the screen.

Approaching the obtained pattern in the subjective cognitive load ratings, at least on a descriptive level, participants reported lower ICL and ECL. On the one hand, these scores might have lacked statistical significance due to the reported insufficient power level of about 71%. On the other hand, as secondary task requirements did not differ between conditions, the absence of considerable differences in ECL is indeed plausible. Moreover, adding just one symbol to the combination might not have resulted in extensive increases in task complexity, potentially explaining the absence of significant differences in ICL. The reversed pattern for GCL could have originated in the particular formulation of the related questions, which emphasizes the subjective impression of understanding and knowledge acquisition. This might have been higher in the easy condition. The authors (Leppink et al., 2013) also discuss this issue in a more recent publication that applies this questionnaire (Leppink, Paas, van Gogh, van der

Vleuten, & Van Merriënboer, 2014). They attribute the lack of meaningful results for GCL to the substantial redundancy between GCL and ICL and take this as further evidence for the initially outlined re-reduction of the three-factor model. By contrast, a more recently developed cognitive load questionnaire by Klepsch, Schmitz, and Seufert (2017) addresses this issue by including the effort component more explicitly in the GCL facet and advocates the existing three-factor model from a measurement perspective. As the authors claim the applicability for a wider range of learning contexts and domains, using this questionnaire instead would be a valuable extension in future studies.

Regarding the schema application task, faster response times in the difficult condition compared to the easy condition hint at higher investments of mental resources with higher levels of task difficulty. This explanation corresponds well with results on contextual interference (de Croock, Van Merriënboer, & Paas, 1998; Van Merriënboer et al., 2006), which report that learners can achieve a more robust and stable transfer performance under conditions that disable fast and easy skill acquisition. Additional support arises from evidence on desirable difficulties in learning situations (Bjork & Bjork, 2011). For instance, when learners must cope with interleaved tasks, they need to maintain sustained engagement of cognitive resources, which fosters performance. The observed increase in response times with increasing distortion may result due to the requirement of additional cognitive operations, as distorted symbols demand the identification of the underlying symbol category before the judgement.

4.1 Implications

The pattern of evidence supports a temporal extension of the CLT framework and reveals a logarithmically decreasing cognitive load progression with increasing schema acquisition. Moreover, results indicate the benefit of an automatic detection of the current level of cognitive load in speech parameters. This could be of value for realizing adaptive user interfaces in digital learning contexts that bear the ability to adjust task complexity and instructional guidance to learners' needs and preferences. A particularly promising field of application comprises foreign language learning, where spoken interaction during the learning process constitutes an essential pre-requisite to shape language skills. User-aligned instructional aids within a considered language training program might then entail additional explanations or calming feedback if states of high load are detected by extended pausing or low articulation frequencies.

4.2 Limitations

A potential limitation arises from the differing symbol complexity, since a visually dominating appearance like a star with salient edges and corners, could foster and speed up schema acquisition and thus benefit symbol recall. More detailed analyses of times spent on drawing different symbols in the retention questionnaire by mouse tracking or gaze pattern analyses might enlighten this issue in future research. With reference to the non-prototypical symbols used for the schema application task, although there was a high level of consensus in the pre-study regarding the correct match to the underlying prototype, subtle differences still could have persisted and influenced the findings in the main task.

The lack of significant differences between conditions in secondary task efficiency could have resulted from task order ambiguities. Although participants were instructed to give equal weight to both tasks, the secondary task occurred first in order and could have been regarded as more easy and familiar. For this reason, participants might have been motivated to prioritize this task instead of the primary task and assign only free resources to the primary task. In consequence, the easy task condition had achieved better results in primary task efficiency, whereas secondary task performance stayed unaffected by task complexity. The choice of keeping this fixed presentation order across all trials and participants aligned to the original task procedure reported for automated complex span tasks (Redick, et al., 2012; Unsworth et al., 2005). Using a counterbalanced presentation order instead could result in more balanced weighting of the priority of both tasks in future studies. A further valuable extension in further studies could address the relationship between both tasks more explicitly to obtain insights into participants' strategies of cognitive resource distribution. The additional use of cognitive models (Anderson, 1983) that compare both task order strategies can clarify distinct effects of task order prioritization.

4.3 Future research

Prospective research should monitor learners' focus on presented symbols or alternative learning material by inspecting gaze behavior in combination with pupil dilation, as suggested by Foroughi, Sibley, and Coyne (2017) and Mitra, McNeal, and Bondell (2017). Another promising extension incorporates the transfer of the obtained patterns and mechanisms to more applied task settings in a different task domain like motor learning. In this domain, an additional step could involve the inclusion of a spatial dimension or the use of animated stimuli or motor sequences, and distinct audio-verbal secondary tasks with a different task order are suitable as well. A further interesting extension takes evidence on modality compatibility in dual-task settings (Hazeltine et al., 2006) into account, and inspects the use of incompatible content-modality matchings.

5 Conclusions

The study applied a multi-measure framework of cognitive load assessment to gain further insights in the temporal progression of cognitive load during schema acquisition. Results replicate the logarithmic pattern of change over the task observed in prior research and reveal influences of task complexity and situational constraints. In summary, the promising approach holds value to address existing research gaps in cognitive load research and gain better insights into changing demands from schema acquisition in learning settings.

Acknowledgements

Thanks to Robert Alig and Laura Seelig for their help with data collection and preparation.

Declarations

Abbreviations

ANOVA: Analysis of variance

CLT: Cognitive Load Theory

ECL: Extraneous cognitive load

GCL: Germane cognitive load

HR: Heart rate

HSD: Honest significant difference

ICL: Intrinsic cognitive load

OSPAN: Operation span task

RMSE: Root mean squared error

SCR: Skin conductance response

SE: Standard error

STEM: Science, Technology, Engineering, and Mathematics

Ethics approval and consent to participate

Informed consent to participate in the study was obtained from each participant in line with the ethical guidelines of the German Society for Psychology (DGPs). The study was approved by the Managing Director of the Institute for Media Research, Faculty of Humanities, Technische Universität (TU) Chemnitz, as formal approval by an ethics committee is not mandatory for psychological research in Germany.

Consent for publication

Not applicable.

Availability of data and material

The materials and the datasets analyzed during the study are available on request.

Competing interests

The authors declare that they have no competing interests.

Funding

The reported research was conducted with grateful funding from the German Research Foundation/DFG grant GRK 1780/1. The publication costs of this article were funded by the German Research Foundation/DFG and TU Chemnitz in the funding program Open Access Publishing.

Authors' contributions

MW is the lead author of the manuscript and the responsible coordinator of the conducted study. MW and RH developed the underlying research design, collected, prepared, and analyzed empirical data, and prepared the initial draft of the manuscript. SEB contributed to the research design, collected and prepared empirical data, and supported data analysis and manuscript preparation. ME and GDR provided critical revisions on the research design and/or the manuscript. All authors read and approved the final manuscript.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA, USA: Harvard University Press.

Atkinson, C. & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*, 82-90. Stable URL: http://www.jstor.org/stable/24922803

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559. doi: 10.1126/science.1736359

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, MA, USA: Cambridge University Press.

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. doi:10.1016/j.learninstruc.2009.02.024

Berthold, A., & Jameson, A. (1999). Interpreting symptoms of cognitive load in speech input. In *UM99 User Modeling* (pp. 235–244). New York, NY, USA: Springer Science + Business Media.

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.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*, 115-132. doi:10.1023/B:TRUC.0000021812.96911.c5

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 109-119. doi:10.1027//1618-3169.49.2.109

Chen, F., Zhou, J., & Yu, K. (2018). Multimodal and data-driven cognitive load measurement. In R. Zheng, (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 3643-4043). New York, NY: Routledge.

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

De Croock, M. B., Van Merriënboer, J. J., & Paas, F. G. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior*, *14*, 249-267. doi:10.1016/S0747-5632(98)00005-3

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

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Mineola, NY, USA: Dover Publications.

Esposito, A., Stejskal, V., Smékal, Z., & Bourbakis, N. (2007). The significance of empty speech pauses: Cognitive and algorithmic issues. In *International Symposium on Brain, Vision, and Artificial Intelligence* (pp. 542–554). Springer.

Foroughi, C. K., Sibley, C., & Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology*, 1-8. doi:10.1111/psyp12896

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema?: A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. doi:10.1016/j.neuropsychologia. 2013.11.010

Hazeltine, E., Ruthruff, E., & Remington, R. W. (2006). The role of input and output modality pairings in dual-task performance: Evidence for content-dependent central interference. *Cognitive Psychology, 52*, 291-345. doi:10.1016/j.cogpsych.205.11.001

Herms, R. (2016). Prediction of deception and sincerity from speech using automatic phone recognition-based features. In N. Morgan (Ed.), *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association* (pp. 2036–2040). San Francisco, CA, USA: ISCA.

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist*, *45*, 1–14. doi:10.1080/00461520903213618

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. doi:10.1007/s10648-010-9150-7

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review*, *28*, 831-852.

Khawaja, M. A., Ruiz, N., & Chen, F. (2007). Potential speech features for cognitive load measurement. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces* (pp. 57–60). ACM.

Khawaja, M. A., Ruiz, N., & Chen, F. (2008). Think before you talk: An empirical study of relationship between speech pauses and cognitive load. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat* (pp. 335–338). ACM.

Klepsch, M., Schmitz, F., & Seufert, T. (2017) Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*:1997. doi: 10.3389/fpsyg.2017.01997.

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review, 30*, 503-529. doi: 10.1007/s10648-017-9404-8

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311-328. doi:10.1037//0021-9010.78.2.311

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19,* 143-148. doi:10.1177/0963721410370295

Lee, K.-F. (1988). *Automatic speech recognition: the development of the SPHINX system* (Vol. 62). New York, NY, USA: Springer Science + Business Media.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45, 1058-1072. doi:10.3758/s13428-013-0334-1

Leppink, J., Paas, F. G. W. C., Van Gog, T., Van der Vleuten, C. P. M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. doi:10.1016/j.learninstruc.2013.12.001

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*, 73-79. doi: 10.1027/1614-2241.4.2.73

Martin, S. (2018). A critical analysis of the theoretical construction and empirical measurement of cognitive load. In R. Zheng (Ed.), *Cognitive load measurement and application: A* *theoretical framework for meaningful research and practice* (pp. 778-1155). New York, NY, USA: Routledge.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. doi:10.3758/s13428-011-0168-7

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics, 5*, 269-287. doi: 10.3102/10769986005003269

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demands from a working memory task: An on-road study across three age groups. *Human Factors, 54*, 396-412. doi:10.1177/0018720812442086

Mitra, R., McNeal, K. S., & Bondell, H. D. (2017). Pupillary response to complex interdependent tasks: A cognitive-load theory perspective. *Behavior Research Methods, 49*, 1905-1919. doi:10.3758/s13428-016-0833-y

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., & Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In *International Conference on User Modeling* (pp. 24–33). New York, NY, USA: Springer Science + Business Media.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.). *Handbook of Perception and Performance. Vol. II. Cognitive Processes and Performance* (pp. 42/1-42/49). New York, NY, USA: Wiley.

O'Keefe, D. J. (2007). Brief report: Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication methods and measures*, *1*, 291-299. doi:10.1080/19312450701641375

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91. doi:10.1016/j.learninstruc.2006.02.004.

Park, B. (2010). *Testing the additivity hypothesis of cognitive load theory* (Doctoral dissertation, Saarland University, Saarbruecken, Germany). Retrieved from https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/23365/1/Dissertation_BP_2010.pdf.

Park, B., & Brünken, R. (2018). Secondary task as a measure of cognitive load. In R. Zheng (Ed.). *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 1901-2317). New York, NY: Routledge.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY, USA: International Universities Press.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, *83*, 157. doi: 10.1037/0033-295X.83.2.157

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164-171. doi:10.1027/1015-5759/a000123

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.

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. doi:10.1037/0033-295X.84.2.127

Sonnenfeld, N. A., & Keebler, J. R. (2016). A quantitative model for unifying human factors with cognitive load theory. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 403-407). Los Angeles, CA, USA: SAGE Publications.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. doi: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., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296. doi:10.1023/A:1022193728205

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. doi:10.3758/BF03192720

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

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. doi:10.1016/j.compedu.2016.04.010

Wirzberger, M., Esmaeili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task. *Acta Psychologica, 179*, 30-41. doi: 10.1016/j.actpsy.2017.07.001

Yap, T. F. (2012). *Speech production under cognitive load: Effects and classification* (Doctoral dissertation, University of New South Wales, Sydney, Australia). Retrieved from http://unsworks.unsw.edu.au/fapi/datastream/unsworks:10507/SOURCE02?view=true

Zheng, R. (2018), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice*. New York, NY, USA: Routledge.

List of tables

Synopsis

Table 1 Cognitive load measures applied by the experimental studies in the included articles 5

Table 2 Sample characteristics of the experimental studies reported in the included articles 10

Table 3 Study designs of the experimental studies reported in the included articles 11

Table 4 Characteristics of data analysis in the experimental studies reported in the included articles 14

Table 5 Parameter settings related to chunk activation and retrieval time 24

Article 1

Table 1 Descriptive values of dependent variables regarding main effects of independent variables 54

Table 2 Standardized beta-coefficients, standard errors, t-values and significance levels of fixed effects in linear mixed model analyses for independent variables 57

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 59

Article 2

Table 1 Comparison of tested conditional growth curve models with linear and/or non-linear predictors 82

Table 2 Descriptive values of learning efficiency in pre-, post- and peri-interruption stages of the task 84

Article 3

Table 1 Categorization and distortion rating of prototypical and distorted symbols 109

Table 2 Descriptive values of correct responses and reaction time for levels of distortion in conditions 120

List of figures

Synopsis

Figure 1 Complexity-related differences in reaction times and errors in both update and final recall stages 13

Figure 2 Overview of ACT-R core modules with corresponding brain regions 19

Figure 3 Outline of steps to perform in each the learning trial of the task 22

Figure 4 Outline of steps to perform in each occurrence of the interrupting task 23

Figure 5 Reaction times for human data and model for the learning task in the easy and difficult condition (correct trials) 25

Figure 6 Accuracy for human data and model for the learning task in the easy and difficult condition 26

Figure 7 Hemodynamic response function (based on SPM) 27

Figure 8 Module activity across different temporal stages of the symbol learning task (excluding resumption trials) 27

Figure 9 Module activity across interruption, resumption, and learning stages of the task 28

Article 1

Figure 1 Sample practice trial sequence for the working memory updating task 50

Figure 2 Experimental manipulations of complexity, split-attention and schema presence 51

Figure 3 Interaction of complexity and schema presence for RT_{recall} 55

Figure 4 Interaction of complexity and split attention for RT_{update} 56

Figure 5 Interaction of complexity, split attention and schema presence for $Errors_{update}$ 58

Article 2

Figure 1 Schematic outline of potential linear and nonlinear temporal models of cognitive load induced during schema acquisition 74

Figure 2 Trial structure within the schema acquisition task 78

Figure 3 Overall changes in efficiency across trials 83

Figure 4 Efficiency in trials immediately before and after an interruption 86

Figure 5 Loss in efficiency due to interruption (pre-post comparison) in easy and difficult conditions 86

Figure 6 Interruption performance over task 88

Article 3

Figure 1 Distorted symbols 108

Figure 2 Schematic outline of the trial procedure in the learning task 112

Figure 3 Stimuli in the schema application task: (a) difficult condition with prototypical symbols and (b) easy condition with distorted symbols 113

Figure 4 Changes in performance in primary task efficiency over the task 116

Figure 5 Changes in performance in secondary task efficiency over the task 117

Figure 6 Observed and fitted logarithmic progressions of speech-related parameters computed per trial in easy (a, c, e) and difficult (b, d, f) conditions 118

Figure 7 Observed and fitted logarithmic progressions of SCR and HR over time in easy (a, c) and difficult (b, d) conditions including both seasonal and trend components 119

Selbstständigkeitserklärung

Vorname, Name: Maria Wirzberger

Geburtsdatum: 15.12.1983

Geburtsort: Lohr am Main, Bayern

Ich erkläre hiermit gegenüber der Technischen Universität Chemnitz, dass ich die vorliegende Dissertation mit dem Titel *Load-inducing factors in instructional design: Process-related advances in theory and assessment* selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Die vorliegende Arbeit ist frei von Plagiaten, der elektronischen Überprüfung auf Plagiate stimme ich zu. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch nicht als Prüfungsleistung für den Erhalt eines akademischen Grades eingereicht.

Ort, Datum Unterschrift

Statement of authorship

I hereby certify to the Technische Universität Chemnitz that this dissertation entitled *Load inducing factors in instructional design – Process-related advances in theory and assessment* is all my own work and uses no external material other than that acknowledged in the text.

This work contains no plagiarism, I agree to an electronical check for plagiarism. All sentences or passages directly quoted from other work or including content derived from such work have been specifically credited to the authors and sources.

This dissertation has not been submitted in the same or similar form for examination related to the award of an academic degree.

Curriculum Vitae

Personal details

Date of Birth: December 15, 1983

Place of Birth: Lohr am Main, Bavaria, Germany

Mail: maria.wirzberger@web.de | *Web:* www.maria-wirzberger.de

Education

04/2015 – 01/2019	**Ph.D. in Psychology (Dr. rer. nat., summa cum laude)**, TU Chemnitz Doctoral thesis: *Load-inducing factors in instructional design: Process-related advances in theory and assessment* Advisors: Prof. Dr. Günter Daniel Rey, Prof. Dr. Josef F. Krems
10/2012 – 12/2014	**Human Factors (Master of Science, Grade A / 1.3)**, TU Berlin Master thesis: *Smart@load? Modeling interruption while using a Smartphone-app in alternating workload conditions* Advisors: Prof. Dr.-Ing. Nele Rußwinkel, Prof. Dr.-Ing. Sebastian Möller
04/2009 – 09/2012	**Psychology (Bachelor of Science, Grade A / 1.3)**, University of Hagen Bachelor thesis: *Experimentelle Induktion beeinträchtigter Aufmerksamkeit im Kontext des seductive detail Effekts* Advisors: Prof. Dr. Günter Daniel Rey, Prof. Dr. Wolfgang Mack
09/2002 – 01/2006	**Inclusive Education (Diploma, Grade A / 1.1)**, Protestant University of Applied Sciences, Bochum Diploma thesis: *"Bruder sein das ist nicht schwer, Schwester sein dagegen sehr?" – Geschlechtsspezifische Betrachtungsweisen zur Situation von Geschwistern behinderter Kinder und Jugendlicher* Advisors: Prof. Dr. Marianne Hellmann, Prof. Dr. Hildegard Mogge-Grotjahn
07/2002	**Fachhochschulreife (Grade A / 1.5)**, Staatliche Fachoberschule und Berufsoberschule Marktheidenfeld

Professional positions

since 11/2018	**Postdoctoral researcher**, Max Planck Research Group "Rationality Enhancement" (Dr. Falk Lieder), Max Planck Institute for Intelligent Systems, Tübingen
04/2018 – 10/2018	**Researcher**, Chair of Psychology of Learning with Digital Media (Prof. Dr. Günter Daniel Rey), TU Chemnitz
04/2017 – 10/2018	**Junior lecturer**, Chair of Psychology of Learning with Digital Media (Prof. Dr. Günter Daniel Rey), TU Chemnitz
09/2018	**Visiting researcher**, Institute of Artificial Intelligence and Cognitive Engineering, Faculty of Science and Engineering, University of Groningen, the Netherlands
04/2015 – 03/2018	**Researcher**, Interdisciplinary Research Training Group "CrossWorlds – Connecting virtual and real social worlds" (funded by the German Research Foundation), TU Chemnitz
02/2018	**Guest lecturer** (Erasmus+ program), Department of Psychology, Faculty of Education, Univerzita Jana Evangelisty Purkyně, Ústí nad Labem, Czech Republic
01/2015 – 03/2015	**Researcher and lecturer**, Chair of Cognitive Modeling in dynamic Human-Machine Systems (Prof. Dr.-Ing. Nele Rußwinkel), TU Berlin
01/2013 – 09/2014	**Tutor**, Chair of Human-Machine Systems (Prof. Dr.-Ing. Matthias Rötting), TU Berlin
09/2010 – 12/2010	**Research intern**, Max Planck Research Group "Affect Across the Lifespan" (Prof. Dr. Michaela Riediger), Max Planck Institute for Human Development, Berlin
04/2009 – 06/2010	**Lecturer**, Vocational Training Center of the Bavarian Economy (bfz), Aschaffenburg
09/2006 – 12/2008	**Pre-primary education professional**, Germany and Switzerland

List of publications

Journal articles with peer-review (including forthcoming articles)

Wirzberger, M., Borst, J. P., Krems, J. F., & Rey, G. D. (in preparation). Peeking behind the curtain: A cognitive modeling approach to exploring memory-related cognitive load effects in an interrupted learning task. *Journal of Memory and Language*.

Wirzberger, M., Esmaeili Bijarsari, S., Herms, R., & Rey, G. D. (in preparation). Gaudy guidance or dreary delusion? Impacts of guidance fading and color coding on schema acquisition in a robot construction task. *Computers & Education*.

Wirzberger, M., Schneider, S., & Rey, G. D. (in preparation). Harm or harness: Do hyperlinks interrupt schema acquisition?. *Computers in Human Behavior*.

Wirzberger, M., Schmidt, R., Georgi, M., Hardt, W., Brunnett, G., & Rey, G. D. (in revision). The impact of system response delay on elderly humans' memory performance in a virtual training scenario. *Scientific Reports*.

Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review*. doi:10.1007/s10648-018-9456-4

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications,* 3:46. doi: 10.1186/s41235-018-0138-z

Schneider, S., Wirzberger, M., & Rey, G. D. (2018). The moderating role of arousal on the seductive detail effect. *Applied Cognitive Psychology*. doi: 10.1002/acp.3473 (Online First)

Wirzberger, M., & Rey, G. D. (2018). Attention please! Enhanced attention control abilities compensate for instructional impairments in multimedia learning. *Journal of Computers in Education, 5*, 243-257. doi: 10.1007/s40692-018-0106-0

Wirzberger, M., Esmaeili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task. *Acta Psychologica, 179,* 30-41. doi: 10.1016/j.actpsy.2017.07.001

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. doi: 10.1016/j.compedu.2016.04.010

Wirzberger, M., & Russwinkel, N. (2015). Modeling interruption and resumption in a smartphone task: An ACT-R approach. *i-com, 14*, 147-154. doi: 10.1515/icom-2015-0033

Russwinkel, N., Prezenski, S., Lindner, S., Halbrügge, M., Schulz, M., & Wirzberger, M. (2014). Modeling of cognitive aspects of mobile interaction. *Cognitive Processing, 15*(Suppl. 1), S22-S24. doi: 10.1007/s10339-014-0632-2

Conference proceedings and book chapters with peer-review (including abstracts)

Wirzberger, M., Schmidt, R., Georgi, M., Hardt, W., Brunnett, G., & Rey, G. D. (2018). Influences of system response delay on elderly participants' performance in a virtual memory training. In R. Wiczorek, D. Manzey, L. Onnasch, K. Brookhuis, A. Toffetti, & D. de Waard (Eds.), *Annual Meeting of the Europe Chapter of the Human Factors and Ergonomics Society 2018, Technology for an Aging Society, Book of Abstracts* (p. 42). Berlin.

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2018). Guidance or Setting? Exploring the learnability of computer-based instructions in a construction task. In J. Hartig, & H. Horz (Eds.), *51st Conference of the German Psychological Society. Abstracts* (p. 509). Lengerich: Pabst Science Publishers.

Herms, R., Wirzberger, M., Eibl, M., & Rey, G. D. (2018). CoLoSS: Cognitive load corpus with speech and performance data from a symbol-digit dual-task. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, … T. Tokunaga (Eds.), *Proceedings of the 11th International Language Resources and Evaluation Conference (LREC 2018)* (pp. 4312-4317). Miyazaki, Japan: European Language Resources Association (ELRA).

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2018). Guidance or Setting? Exploring the learnability of computer-based instructions in a construction task. In A. C. Schütz, A. Schubö, D. Endres, & H. Lachnit (Eds.), *Abstracts of the 60th Conference of Experimental Psychologists* (p. 69). Lengerich: Pabst Science Publishers.

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Rey, G. D., & Eibl, M. (2018). Cognitive load influences performance, speech and physiological parameters in a multimodal dual-task setting. In A. C. Schütz, A. Schubö, D. Endres, & H. Lachnit (Eds.), *Abstracts of the 60th Conference of Experimental Psychologists* (p. 295). Lengerich: Pabst Science Publishers.

Wirzberger, M., Schmidt, R., Rey, G. D., & Hardt, W. (2017). Auswirkung systeminduzierter Delays auf die menschliche Gedächtnisleistung in einem virtuellen agentenbasierten Trainingssetting. In M. Eibl, & M. Gaedke (Eds.), *INFORMATIK 2017, Lecture Notes in*

Informatics (LNI) (pp. 2287-2294). Bonn: Gesellschaft für Informatik. doi: 10.18420/in2017_229

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2017). Lernförderliche Gestaltung computerbasierter Instruktionen zur Roboterkonstruktion. In M. Eibl, & M. Gaedke (Eds.), *INFORMATIK 2017, Lecture Notes in Informatics (LNI)* (pp. 2279-2286). Bonn: Gesellschaft für Informatik. doi: 10.18420/in2017_228

Wirzberger, M., Truschzinski, M., Schmidt, R., & Barlag, M. (2017). Computer Science meets Cognition: Möglichkeiten und Herausforderungen interdisziplinärer Kognitionsforschung. In M. Eibl, & M. Gaedke (Eds.), *INFORMATIK 2017, Lecture Notes in Informatics (LNI)* (pp. 2273-2277). Bonn: Gesellschaft für Informatik. doi: 10.18420/in2017_227

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Rey, G. D., & Eibl, M. (2017). Influences of cognitive load on learning performance, speech and physiological parameters in a dual-task setting. In *Abstracts of the 20th Conference of the European Society for Cognitive Psychology* (p. 161). Potsdam.

Wirzberger, M., Rey, G. D., & Krems, J. F. (2017). Modeling cognitive load effects in an interrupted learning task: An ACT-R approach. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 3540-3545). Austin, TX: Cognitive Science Society.

Truschzinski, M., & Wirzberger, M. (2017). A dynamic process model for predicting workload in an air traffic controller task. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1224-1229). Austin, TX: Cognitive Science Society.

Wirzberger, M. (2017). Inspecting cognitive load factors in digital learning settings with ACT-R. In T. Arnold, B. Heinzerling, C. Tauchmann, T. Grube, M. Maaß, & N. Wessels (Eds), *Dagstuhl 2017. Proceedings of the 11th Joint Workshop of the German Research Training Groups in Computer Science* (p. 62).

Wirzberger, M., Schneider, S., Dlouhy, S., & Rey, G. D. (2017). Time – Space – Content? Interrupting features of hyperlinks in multimedia learning. In T. Goschke, A. Bolte, & C. Kirschbaum (Eds.), *Abstracts of the 59th Conference of Experimental Psychologists (TeaP)* (p. 97). Lengerich: Papst Science Publishers.

Schneider, S., Wirzberger, M., Augustin, Y., & Rey, G. D. (2017). The moderating role of arousal on the seductive detail effect. In T. Goschke, A. Bolte, & C. Kirschbaum (Eds.), *Abstracts of the 59th Conference of Experimental Psychologists (TeaP)* (p. 96). Lengerich: Papst Science Publishers.

Beege, M., Nebel, S., Schneider, S., Wirzberger, M., Schmidt, N., & Rey, G. D. (2016). Bedingt räumliche Nähe bessere Lernergebnisse? Die Rolle der Distanz und Integration beim Lernen mit multiplen Informationsquellen. In I. Fritsche (Ed.), *50th Conference of the German Psychological Society. Abstracts* (p. 540). Lengerich: Pabst Science Publishers

Wirzberger, M., & Rey, G. D. (2016). Examining load-inducing factors in instructional design: An ACT-R approach. In D. Reitter, & F. E. Ritter (Eds.), *Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016)* (pp. 223-224). University Park, PA: Penn State.

Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016). CLT meets WMU: Simultaneous experimental manipulation of load factors in a basal working memory task. In *9th International Cognitive Load Theory Conference, June 22nd to 24th, 2016, Bochum, Germany, Abstracts* (p. 19).

Wirzberger, M. (2016). Modeling load factors in multimedia learning: An ACT-R approach. In B. Etzold, R. Richter, M. Eibl, & W. Lehner (Eds.), *Dagstuhl 2016. Proceedings of the 10th Joint Workshop of the German Research Training Groups in Computer Science* (p. 98). Chemnitz: Universitätsverlag Chemnitz.

Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016). Separating cognitive load facets in a working memory updating task: An experimental approach. In *International Meeting of the Psychonomic Society, Granada – Spain, May 5-8, 2016, Abstract Book* (pp. 211-212).

Wirzberger, M., & Rey, G. D. (2016). CLT meets ACT-R: Modeling load-inducing factors in instructional design. In J. Funke, J. Rummel, & A. Voß (Eds.), *Abstracts of the 58th Conference of Experimental Psychologists* (p. 377). Lengerich: Pabst Science Publishers.

Lüderitz, C., Wirzberger, M., & Karrer-Gauß, K. (2016). Sustainable effects of simulator-based training on ecological driving. In B. Deml, P. Stock, R. Bruder, & C. M. Schlick (Eds.), *Advances in Ergonomic Design of Systems, Products and Processes. Proceedings of the Annual Meeting of the GfA 2015* (pp. 463-475), Berlin, Heidelberg: Springer.

Wirzberger, M. & Rey, G. D. (2015). Cognitive modeling meets instructional design: Exploring Cognitive Load Theory with ACT-R. In C. Wienrich, T. Zander, & K. Gramann (Eds.), *Trends in Neuroergonomics. Proceedings of the 11th Berlin Workshop Human-Machine*

Systems (pp. 190-193), Berlin: Universitätsverlag der TU Berlin. doi: 10.14279/depositonce-4887

Lüderitz, C., Wirzberger, M., & Karrer-Gauß, K. (2015). Nachhaltige Effekte simulatorbasierten Trainings auf eine ökologische Fahrweise. In Gesellschaft für Arbeitswissenschaft e.V. (Ed.), *VerANTWORTung für die Arbeit der Zukunft, 61st Conference of the Society for Ergonomics and Work Science, Karlsruhe Institute of Technology (KIT), Institute of Human and Industrial Engineering (ifab), February 25th-27th, 2015*, Dortmund: GfA Press.

Wirzberger, M., Lüderitz, C., Rohrer, S., & Karrer-Gauß, K. (2014). „Keep green!" – Nachhaltige Förderung ökologischen Fahrens durch Simulatortraining?. In Güntürkün, O. (Ed.), *49th Conference of the German Psychological Society. Abstracts* (p. 570), Lengerich: Pabst Science Publishers.

Wirzberger, M., & Russwinkel, N. (2014). "I don't need it" – Modeling ad-induced interruption while using a smartphone-app. In *CrossWorlds 2014: Theory, Development and Evaluation of Social Technology*, Chemnitz. doi: 10.13140/RG.2.1.4426.0966

Wirzberger, M., & Rey, G. D. (2013). Attention impairment in multimedia learning: Does initial task attention act as moderator?. In Schwab, F., Carolus, A., Brill, M., & Hennighausen, C. (Eds.), *Media Psychology: Media Research: Yesterday, Today, and Tomorrow. Proceedings of the 8th Conference of the Media Psychology Division of the German Psychological Society* (p. 11), Würzburg: Universität Würzburg.

Wirzberger, M., & Rey, G. D. (2013). Inducing impaired attention within the seductive detail effect: Do already distracted learners suffer more?. In Ansorge, U., Kirchler, E., & Lamm, C., & Leder, H. (Eds.), *Abstracts of the 55th Conference of Experimental Psychologists* (p. 314), Lengerich: Pabst Science Publishers.

Conference contributions (peer-review), workshops and invited talks

Wirzberger, M. (2018, October 30). *Pädagogik trifft Psychologie und Informatik: Interdisziplinäre Perspektiven zur Gestaltung intelligenter nutzer- und kontextadaptiver Lehr-Lernsysteme*. Invited talk at the University of Stuttgart.

Wirzberger, M., Schmidt, R., Georgi, M., Hardt, W., Brunnett, G., & Rey, G. D. (2018, October, 8). *Influences of system response delay on elderly participants' performance in a virtual memory training*. Poster at the Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Berlin.

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2018, September 19). *Guidance or Setting? Exploring the learnability of computer-based instructions in a construction task*. Poster at the 51st Congress of the German Psychological Society (DGPs), Frankfurt.

Wirzberger, M. (2018, September 14). *Modeling interruption effects in an abstract learning task*. Invited talk at the Cognitive Modeling Group at the University of Groningen.

Wirzberger, M. (2018, August 24). *Prozessbezogene kognitive Beanspruchung in digitalen Lernszenarien: Eine grundlagenorientierte Betrachtung*. Invited talk at the Cognitive Systems Group at the University of Bamberg.

Wirzberger, M. (2018, August 16). *Designing intelligent educational technologies: Perspectives related to distraction, interruption, and cognitive load*. Invited talk at the Rationality Enhancement Group at the Max Planck Institute for Intelligent Systems, Tübingen.

Herms, R., Wirzberger, M., Eibl, M., & Rey, G. D. (2018, May 11). *CoLoSS: Cognitive load corpus with speech and performance data from a symbol-digit dual-task*. Poster at the 11th International Language Resources and Evaluation Conference (LREC 2018), Miyazaki, Japan.

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2018, March 12). *Guidance or Setting? Exploring the learnability of computer-based instructions in a construction task*. Poster at the 60th Conference of Experimental Psychologists (TeaP), Marburg.

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Rey, G. D., & Eibl, M. (2018, March 13). *Cognitive load influences performance, speech and physiological parameters in a multimodal dual-task setting*. Talk at the 60th Conference of Experimental Psychologists (TeaP), Marburg.

Wirzberger, M., Schmidt, R., Rey, G. D., & Hardt, W. (2017, September 25). *Auswirkung systeminduzierter Delays auf die menschliche Gedächtnisleistung in einem virtuellen agentenbasierten Trainingssetting*. Talk at the INFORMATIK 2017, Chemnitz.

Esmaeili Bijarsari, S., Wirzberger, M., & Rey, G. D. (2017, September 25). *Lernförderliche Gestaltung computerbasierter Instruktionen zur Roboterkonstruktion*. Talk at the INFORMATIK 2017, Chemnitz.

Wirzberger, M., Herms, R., Esmaeili Bijarsari, S., Rey, G. D., & Eibl, M. (2017, September 4). *Influences of cognitive load on learning performance, speech and physiological parameters in a dual-task setting*. Poster at the 20th Conference of the European Society for Cognitive Psychology, Potsdam.

Wirzberger, M., Rey, G. D., & Krems, J. F. (2017, July 29). *Modeling cognitive load effects in an interrupted learning task: An ACT-R approach*. Poster at the 39th Annual Meeting of the Cognitive Science Society (CogSci), London, UK.

Truschzinski, M., & Wirzberger, M. (2017, July 27). *A dynamic process model for predicting workload in an air traffic controller task.* Talk at the 39th Annual Meeting of the Cognitive Science Society (CogSci), London, UK.

Wirzberger, M. (2017, June 7). *CLT meets ACT-R: Modellierung von Facetten kognitiver Beanspruchung in Lernsituationen*. Invited talk at the Chair of Cognitive Psychology & Cognitive Ergonomics at TU Berlin.

Wirzberger, M., Schneider, S., Dlouhy, S., & Rey, G. D. (2017, March 27). *Time – Space – Content? Interrupting features of hyperlinks in multimedia learning*. Talk at the 59th Conference of Experimental Psychologists (TeaP), Dresden.

Schneider, S., Wirzberger, M., Augustin, Y., & Rey, G. D. (2017, March 27th). *The moderating role of arousal on the seductive detail effect*. Talk at the 59th Conference of Experimental Psychologists (TeaP), Dresden.

Beege, M., Nebel, S., Schneider, S., Wirzberger, M., Schmidt, N., & Rey, G. D. (2016, September 21). *Bedingt räumliche Nähe bessere Lernergebnisse? Die Rolle der Distanz und Integration beim Lernen mit multiplen Informationsquellen*. Talk at the 50th Conference of the German Psychological Society, Leipzig.

Wirzberger, M., & Rey, G. D. (2016, August 4). *Examining load-inducing factors in instructional design: An ACT-R approach*. Poster at the 14th International Conference on Cognitive Modeling, University Park, Pennsylvania, USA.

Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016, June 22). *CLT meets WMU: Simultaneous experimental manipulation of load factors in a basal working memory task*. Poster at the 9th International Cognitive Load Theory Conference, Bochum.

Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016, May 7). *Separating cognitive load facets in a working memory updating task: An experimental approach*. Poster at the International Meeting of the Psychonomic Society, Granada, Spain.

Beege, M., Schneider, S., Nebel, S., Wirzberger, M., Rey, G. D. (2016, March 22). *Look into my eyes! Exploring the effect of addressing in multimedia learning*. Talk at the 58th Conference of Experimental Psychologists (TeaP), Heidelberg.

Nebel, S., Schneider, S., Beege, M., Wirzberger, M., Rey, G. D. (2016, March 22). *Using the jigsaw principle to increase task interdependence in cooperative educational videogames*. Talk at the 58th Conference of Experimental Psychologists (TeaP), Heidelberg.

Wirzberger, M., & Rey, G. D. (2016, March 21). *CLT meets ACT-R: Modeling load-inducing factors in instructional design*. Talk at the 58th Conference of Experimental Psychologists (TeaP), Heidelberg.

Wirzberger, M., & Rey, G. D. (2015, October 9). *Cognitive modeling meets instructional design: Exploring Cognitive Load Theory with ACT-R*. Talk at the 11th Berlin Human-Machine Systems Workshop, Berlin.

Wirzberger, M. (2015, July 24). *CLT meets ACT-R: Modellierung kognitiver Prozesse in einem virtuellen kollaborativen Lernszenario*. Talk at the 2nd Summer School Human Factors, TU Berlin.

Lüderitz, C., Wirzberger, M., & Karrer-Gauß, K. (2015, February 27). *Nachhaltige Effekte simulatorbasierten Trainings auf eine ökologische Fahrweise*. Talk at the 61st Conference of the Society for Ergonomics and Work Science, Karlsruhe.

Wirzberger, M., Lüderitz, C., Rohrer, S., & Karrer-Gauß, K. (2014, September 25). *„Keep green!" – Nachhaltige Förderung ökologischen Fahrens durch Simulatortraining?*. Poster at the 49th Conference of the German Psychological Society, Bochum.

Wirzberger, M., & Russwinkel, N. (2014, July 1). *„I don't need it" – Modeling ad-induced interruption while using a smartphone-app*. Talk at the Crosswolds 2014: Theory, Development and Evaluation of Social Technology, Chemnitz.

Wirzberger, M. (2014, April 11). *Interruption by product ads while using a shopping-app in different workload conditions*. Talk at the 4th ACT-R Spring School and Master Class, Groningen.

Wirzberger, M., & Rey, G. D. (2013, September 5). *Attention impairment in multimedia learning: Does initial task attention act as moderator?*. Talk at the 8th Conference of the Media Psychology Division of the German Psychological Society, Würzburg.

Wirzberger, M., & Rey, G. D. (2013, March 27). *Inducing impaired attention within the seductive detail effect: Do already distracted learners suffer more?*. Talk at the 55th Conference of Experimental Psychologists (TeaP), Vienna.