



Fostering Deep Learning in Immersive Virtual Reality: Interplay of Learner's Characteristics with Internal and External Support

Cumulative dissertation for the degree of

– *Doctor rerum naturalium (Dr.rer.nat.)* –

presented to the Faculty of Engineering, Computer Science and Psychology at Ulm University

Andrea Gwendolin Vogt

from Stuttgart

2021

Department of Learning and Instruction

Institute of Psychology and Education

Acting Dean: Prof. Anke Huckauf
First Referee: Prof. Tina Seufert
Second Referee: Prof. Steffi Zander
Third Referee: Prof. Detlev Leutner
Date of Disputation: 09.05.2022



Over every mountain, there is a path, although it may not be seen from the valley.

(Theodore Roethke)

Acknowledgments

Empirical research is always a special adventure full of unexpected challenges, ups, and downs, which I fortunately did not have to face alone. Many people have supported me throughout this adventure and the creation of the final manuscript, and without their support, this thesis would not have been possible. Firstly, I would like to thank Prof. Dr. Tina Seufert for her constructive discussions and individual feedback that enabled me to constantly learn, develop, and grow as a researcher. Many thanks to Prof. Dr. Steffi Zander for the enriching cooperation in the AuCity2 project, the helpful feedback on my empirical studies, and for reviewing my thesis. I would like to express my special thanks to Prof. Dr. Martin Baumann, who encouraged me to write my doctoral thesis and who repeatedly enhanced my work by asking critical questions.

My co-authors also played a central role in the conception of my work. Without them, these studies would not have been possible. A very special thanks goes to Thomas Witte and Franziska Babel. In particular, I would like to thank my (former) colleagues Lisa Respondek, Nadja Gentner, Melina Klepsch, and Valentin Riemer for sharing their expertise and encouragement.

I would also like to thank all of my colleagues, as well as the research assistants, secretaries, and assistants at the Department of Learning and Instruction who supported me during my doctoral studies. Last but not least, I would like to thank my family and my friends for their encouragement and emotional support.

List of scientific publications

Peer-reviewed articles included in this dissertation

This thesis is based on the three following peer-reviewed journal articles. The theories, methods, results, and discussion points presented in this thesis have already been partially published in these articles.

Vogt, A., Albus, P., & Seufert, T. (2021). Learning in virtual reality: Bridging the motivation gap by adding annotations. *Frontiers in Psychology*, 12, 645032.
<https://doi.org/10.3389/fpsyg.2021.645032>

Journal Metrics: 2-Year Impact Factor: 2.990 (2020), h-index 110

This paper was published in Frontiers in Psychology under the terms of the Creative Commons CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model. *Computers & Education*, 171, 104235.
<https://doi.org/10.1016/j.compedu.2021.104235>

Journal Metrics: 2-Year Impact Factor: 8.538 (2020), h-index 179

This paper was published in the Journal Computers & Education under the terms of the Creative Commons CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome. *British Journal of Educational Technology*, 00, 1–19.
<https://doi.org/10.1111/bjet.13104>

Journal Metrics: 2-Year Impact Factor: 3.241 (2020), h-index 95

This paper was published in the British Journal of Educational Technology under the terms of the Creative Commons CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Further peer-reviewed publications

Journal articles

- Albus, P., **Vogt, A.**, & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, 166, Article 104154. <https://doi.org/10.1016/j.compedu.2021.104154>
- Vogt, A.**, Klepsch, M., & Seufert, T. (2020). Learning from multiple representations: Prior knowledge moderates the beneficial effects of signals and abstract graphics. *Frontiers in Psychology*, 11, Article 601125. <https://doi.org/10.3389/fpsyg.2020.601125>

Conference publications

2021

Vogt, A., Albus, P., Montag, M., Drey, T., Hartwig, S., Tasliarmut, F., & Wolf, M. (2021, September 14-16). *Effekte von Segmentierung und Color Coding in nicht-immersiven Virtual-Reality-Lernumgebungen* [Effects of segmentation and color coding in non-immersive virtual reality learning environments][Symposium]. 18. Fachgruppentagung der Fachgruppe Pädagogische Psychologie (PAEPSY), Heidelberg/online, Germany.

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021, August 23–27). *Learning robotics in virtual reality: Impact of sequencing and prompting on learning outcome*. 19th Biennial EARLI Conference “Education and Citizenship: Learning and Instruction and the Shaping of Futures”, Gothenburg/online, Västra Götaland, Sweden.

Vogt, A., Witte, T., Tichy, M., & Seufert, T. (2021, August 23–27). *Supportive elements for coherence formation increase code comprehension*. 19th Biennial EARLI Conference “Education and Citizenship: Learning and Instruction and the Shaping of Futures”, Gothenburg/online, Västra Götaland, Sweden.

Albus, P., **Vogt, A.**, & Seufert, T. (2021, August 23–27). *Modality effect in virtual reality*. 19th Biennial EARLI Conference “Education and Citizenship: Learning and Instruction and the Shaping of Futures”, Gothenburg/online, Västra Götaland, Sweden.

Vogt, A., Witte, T., Tichy, M., & Seufert, T. (2021, March 14–16). *Impact of a bidirectional 3D code preview on task performance and learning outcome*. Conference of Experimental Psychologists: TeaP@Home, Ulm/online, Baden-Württemberg, Germany.

Krischler, J., **Vogt, A.**, Albus, P., & Koch, C. (2021). *Design principles affecting motivational and cognitive requirements for VR learning environments in engineering education*. In E. Toledo Santos & S. Scheer (Eds.), Proceedings of the 18th International Conference on Computing in Civil and Building Engineering: ICCCBE 2020 (pp. 1175-1186). Springer Nature.

2020

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2020, August 24–25). *Learning about robotics in virtual reality*. EARLI SIG 6 & 7 Conference, Tübingen/online, Baden-Württemberg, Germany.

Albus, P., **Vogt, A.**, & Seufert, T. (2020, August 24–25). *Virtual reality in classrooms*. EARLI SIG 6 & 7 Conference, Tübingen/online, Baden-Württemberg, Germany.

Vogt, A., Albus, P., Schüle, M., & Seufert, T. (2020, March 25–27). *Lernen in der virtuellen Realität: Wie beeinflussen Motivation und Gestaltungsvarianten den Lernerfolg?* [Learning in virtual reality: How does motivation and design options influence learning effort?]. 8th GEBF Conference “Bildung gestalten - Partizipation erreichen - Digitalisierung nutzen”, Potsdam, Brandenburg, Germany. (Conference moved to digiGEBF 2021)

Albus, P., **Vogt, A.**, & Seufert, T. (2020, March 25–27). *Lernen der Zukunft – Welchen Einfluss haben Gestaltungsprinzipien in Virtual Reality auf Lernerfolg und Cognitive Load* [Learning in the future – Which influence will design principles in virtual reality have on learning effort and cognitive load] [Symposium]. 8th GEBF Conference “Bildung gestalten - Partizipation erreichen - Digitalisierung nutzen”, Potsdam, Brandenburg, Germany. (Conference moved to digiGEBF 2021)

2019

Vogt, A., & Seufert, T. (2019, August 12–16). *Fostering learning strategies to support self-regulated learning: A multilevel analysis*. 18th Biennial EARLI Conference “Thinking Tomorrow’s Education: Learning from the past, in the present and for the future”, Aachen, North Rine-Westphalia, Germany.

Abstract

Virtual Reality Learning Environments (VRLEs) refer to computer-generated scenarios that offer great potential for providing learning content. However, VRLEs can also pose challenges to learners regarding visual complexity, integration of different presentation formats, and choosing adequate learning behavior. As a result, such VRLE-related challenges can lead to reduced learning outcomes.

The focus of this thesis was to examine different support approaches (external and internal) and their effects on cognitive learning processes and learning outcomes. Within this, the *Extended Model of Immersive Learning in VR* (EMIL VR) was developed to describe the underlying learning processes and relevant influencing factors on learning outcomes. The postulated EMIL VR was examined in the present thesis based on three empirical studies comparing different support approaches with a control group while learning in VR.

In the first study, the VRLE contained annotations as external support and was compared to a VRLE without annotations. This support approach aimed to reduce visual search and foster mapping of information across different presentation formats. Intrinsic motivation moderated the beneficial effect of annotations on the overall learning outcome.

In the second study, an elaboration prompt functioned as internal support and activated learners to process learning content on a deeper level. The control group did not receive the activating prompt before entering the VRLE. The study included different measurement approaches of learning outcomes: functional and structural. Prompting had beneficial effects on the functional learning outcome (knowledge level). Moreover, learners with low structural prior knowledge also experienced beneficial effects from prompting.

In the third study, external and internal support approaches were combined to investigate possible synergic effects of adequate sequencing and prompting. Each support approach had different effects on the learning subprocesses. Specifically, prompting supported deep learning while adequate sequencing eased superficial processing of the learning content. Thus, no synergic effects were found when combining both support approaches.

Throughout the present thesis, empirical findings will be discussed based on the postulated EMIL VR framework, and additional influencing factors will be critically reflected. All findings of the three studies underline the necessity of supporting learners in VRLEs. Considering the experimental results, theoretical, methodological, and practical implementations are provided.

Zusammenfassung

Virtual-Reality-Lernumgebungen (VRLEs) bezeichnen computergenerierte Szenarien, die ein großes Potenzial für die Vermittlung von Lerninhalten aufweisen. Die Darstellung der Lerninhalte in VRLEs geht jedoch mit potenziellen Herausforderungen für die Lernenden einher: hohe visuelle Komplexität, Integration verschiedener Präsentationsformate sowie die Wahl des geeigneten Lernverhaltens. Diese Herausforderungen in VRLEs können den Lernerfolg beeinträchtigen. Der Schwerpunkt dieser Arbeit lag auf der experimentellen Untersuchung verschiedener Unterstützungsansätze (extern und intern) und ihren Auswirkungen auf kognitive Lernprozesse sowie den Lernerfolg. Zur Beschreibung der zugrunde liegenden Lernprozesse und relevanten Einflussfaktoren wurde das *Erweiterte Modell des immersiven Lernens in VR* (*Extended Model of Immersive Learning in VR*; EMIL VR), entwickelt. Das postulierte Modell, EMIL VR, wurde anhand von drei empirischen Studien geprüft, die unterschiedliche Unterstützungsansätze beim Lernen in VR mit einer Kontrollgruppe verglichen:

In der ersten Studie wurde die VRLE mit Annotationen als externe Unterstützung versehen und mit einer VRLE ohne Annotationen verglichen. Dieser Unterstützungsansatz zielte darauf ab, die visuelle Suche zu reduzieren und die Zuordnung von Informationen über verschiedene Präsentationsformate hinweg zu erleichtern. Intrinsische Motivation moderierte den positiven Effekt der Annotationen auf den Gesamtlernerfolg.

In der zweiten Studie erhielten die Lernenden einen Elaborationsprompt, welcher eine Form der internen Unterstützung darstellte. Der Elaborationsprompt sollte Lernende dazu anregen, den Lerninhalt tiefer zu verarbeiten. Die Kontrollgruppe erhielt vor Beginn der VRLE keine solche aktivierende Komponente. Die Studie umfasste verschiedene Messansätze, um den Lernerfolg zu erfassen: funktional und strukturell. Der Elaborationsprompt wirkte sich positiv auf den funktionalen Lernerfolg auf der Wissensebene aus. Bei Lernenden mit geringem strukturellen Vorwissen wirkte sich der Elaborationsprompt ebenfalls positiv aus.

In der dritten Studie wurden externe und interne Unterstützungsansätze kombiniert, um die Synergieeffekte von adäquater Sequenzierung und Hinzufügen des Elaborationsprompts zu untersuchen. Jeder Unterstützungsansatz zeigte unterschiedliche Auswirkungen auf die Teilprozesse beim Lernen. Der Elaborationsprompt förderte das semantische Verarbeiten der Lerninhalte, während adäquate Sequenzierung die oberflächliche Verarbeitung der Lerninhalte erleichterte. Bei der Kombination beider Unterstützungsansätze konnte kein Synergieeffekt festgestellt werden.

Die empirischen Befunde werden auf der Basis des Rahmenmodells EMIL VR diskutiert und zusätzliche Einflussfaktoren kritisch reflektiert. Die Ergebnisse aller drei Studien unterstreichen die Notwendigkeit der Unterstützung von Lernenden in VRLEs. Basierend auf den empirischen Ergebnissen werden theoretische, methodische und praktische Implikationen abgeleitet.

Contents

Acknowledgments	iii
List of scientific publications	v
Abstract	ix
Zusammenfassung	xi
I. Introduction, theoretical background, related work, and research aims	1
1. Introduction	3
2. Theoretical background and related work	5
2.1. Virtual reality learning environments	5
2.1.1. Defining virtual reality environments	5
2.1.2. Specific qualities of virtual reality learning environments	6
2.2. Empirical findings on learning in virtual reality	8
2.2.1. Cognitive Affective Model of Immersive Learning	9
2.2.2. Challenges while learning in virtual reality	11
2.2.3. Fostering deep learning in virtual reality	12
2.3. Relevant factors for fostering learning in virtual reality learning environments .	13
2.4. Underlying cognitive learning processes	15
2.4.1. Cognitive Theory of Multimedia Learning	15
2.4.2. Integrated Model of Text and Picture Comprehension	16
2.5. Designing appropriate learning material	18
2.5.1. Adding annotations in virtual reality	19
2.5.2. Sequencing textual and pictorial representations	20
2.6. Active role of learners	21
2.6.1. Activating learners by cognitive prompts	22
2.7. Influence of learner's characteristics	23
2.7.1. Intrinsic motivation	24
2.7.2. Prior knowledge	25
2.7.3. Further characteristics of learners	26
2.8. Assessment of learning outcomes	27
2.8.1. Nature of mental models	27

2.8.2. Stages of development of mental models	28
2.8.3. Measuring mental models	28
3. Research aims of the present thesis	31
3.1. Extended Model of Immersive Learning in Virtual Reality	31
3.1.1. Cognitive processes while learning in virtual reality	31
3.1.2. External and internal support approaches for learning in virtual reality	32
3.1.3. Critical learner's characteristics for learning in virtual reality	34
3.1.4. Assessing learning outcomes after learning in virtual reality	35
II. Empirical research	39
4. Overview of the three included studies	41
4.1. Study 1 - Learning in Virtual Reality	44
4.1.1. Theoretical background	44
4.1.2. Research questions and hypotheses	44
4.1.3. Method and results	45
4.1.4. Discussion	45
4.2. Study 2 - Prompting in-depth learning in immersive virtual reality	47
4.2.1. Theoretical background	47
4.2.2. Research questions and hypotheses	47
4.2.3. Method and results	48
4.2.4. Discussion	48
4.3. Study 3 - Immersive virtual reality or auditory text first?	50
4.3.1. Theoretical background	50
4.3.2. Research questions and hypotheses	50
4.3.3. Method and results	51
4.3.4. Discussion	51
III. Discussion and implications	53
5. Discussion	55
5.1. Main findings of the present thesis	55
5.1.1. External support approach for selecting and integrating learning content	56
5.1.2. Prompting deep learning in virtual reality learning environments	59
5.1.3. Combining internal and external support for learning in virtual reality .	61
5.1.4. Learner's characteristics and learning content	64
5.1.5. Assessing learning outcomes	65
5.1.6. Implications of the present thesis	67
5.1.7. Limitations and further recommendations for future studies	68
5.2. Conclusion	72

References	73
IV. Appendix	87
Appendix A - Peer-reviewed journal articles	89
Appendix B - Additional analysis	147
Appendix C - Exemplary R code	155
List of Figures	187
List of Acronyms	189
Declaration	191

Part I.

**Introduction, theoretical background,
related work, and research aims**

1. Introduction

An increasing research interest in Virtual Reality Learning Environments (VRLEs) has become apparent in recent years (Makransky & Petersen, 2021). VRLEs are characterized as 3D computer simulations that are interactive, multisensory, and provide an opportunity for situated learning in many fields (Baceviciute, Mottelson, Terkildsen, & Makransky, 2020). Learners describe their experiences with VRLEs as joyful and motivating (Makransky, Andreassen, Baceviciute, & Mayer, 2020), which speaks to the potential of VRLEs as contemporary learning environments that can be implemented within universities (Miller, Castillo, Medoff, & Hardy, 2021). Despite the theoretical promise that learning in VR holds, empirical studies examining the potential of these learning environments show heterogeneous effects (Wu, Yu, & Gu, 2020). While some studies have found beneficial effects of learning in VR, others have documented superficial processing of learning content and lower learning outcomes compared to conventional learning material (Makransky et al., 2020; Parong & Mayer, 2020).

Adding to this, prior research has identified different challenges that can arise from learning in VR, such as high visual complexity (Ragan et al., 2015; Mulders, Buchner, & Kerres, 2020), lack of adequate learning behavior (Hu-Au & Okita, 2021; Makransky et al., 2020; Huang, Luo, Yang, Lu, & Chen, 2020), as well as inadequate strategies and potential overload while learning (Makransky & Petersen, 2021; Mayer, 2020). To overcome these learning barriers, different approaches aimed at supporting learners in VR have been examined and have found that some support approaches are more effective in VR than in conventional settings (Klingenberg et al., 2020). This implies that support approaches cannot simply be transferred from conventional settings to VR (Meyer, Omdahl, & Makransky, 2019). However, until now, the question of how to effectively support learning in VR remains unanswered. In addition, no empirically investigated frameworks exist to date for describing underlying cognitive processes and relevant influencing factors of successful learning outcomes in VR.

To address these research gaps, the aim of the present thesis was to develop a theoretical framework for learning in VR with a specific focus on different support approaches and their consequences for underlying cognitive learning processes: the *Extended Model for Immersive Learning* in VR (EMIL VR). Additionally, this model considers the interplay of support approaches and learner's characteristics.

Following the recommendation of Jenkins (1979), the present thesis focused on four specific factors that can impact underlying cognitive learning processes. These factors included the learning setting (F1), learning material (F2), learner's characteristics (F3), and assessment of learning outcomes (F4). Based on the deduced challenges of learning in VR, different support approaches were defined based on prominent theories and prior findings. External support

1. Introduction

approaches related to the learning material (F2) as well as internal support approaches related to the learning setting and the active role of learners (F1) were systematically manipulated. Relevant learner's characteristics (F3) were additionally examined, as well as their impact on the effectiveness of the different support approaches. Moreover, different approaches to assess learning outcomes (F4) were included to gain deeper insights into the underlying subprocesses that take place while learning in VR. The results of the empirical studies are critically reflected on and discussed based on the postulated theoretical EMIL VR framework. In conclusion, theoretical, methodological, and practical implications are outlined, important limitations are described, and recommendations for future research are provided.

2. Theoretical background and related work

Learning content can be presented and combined in VR in many ways. To understand the potential of VR as a learning environment, an in-depth overview will first be provided (Section 2.1). Based on the described characteristics of VRLEs, a model for learning in VR, which is based on previous empirical findings, will be comprehensively discussed. Next, the need for learning support within VR will be emphasized and explained in line with prior research findings (Section 2.2). The importance of the four critical factors, namely the learning setting (F1), learning material (F2), learner's characteristics (F3), and assessment of learning outcomes (F4) will be discussed to systematically investigate different support approaches for learning in VR (Section 2.3). To understand the fundamental mechanisms of how different support approaches work, the underlying cognitive learning processes will be outlined (Section 2.4). Subsequently, different support approaches will be analyzed based on previous empirical and theoretical findings. To this end, external support through the design of the VRLE as well as internal support through the activation of appropriate learner behavior will be the focus (Sections 2.5 and 2.6). In particular, it is shown that learner characteristics play an important role in the extent to which support approaches can have a positive effect on learning outcomes (Section 2.7). To gather comprehensive insights into the impact of support in VR, it is important that learning outcomes are considered and measured in a differentiated manner. Thus, the measurement of learning outcomes will be examined from two perspectives: functional and structural (Section 2.8). The advantage of using a multi-method approach will thereby become transparent. At the end of this chapter, the guiding questions of the thesis will be presented, and the overarching theoretical framework model (EMIL VR) will be built up step by step and linked to the leading questions (Chapter 3).

2.1. Virtual reality learning environments

To understand what makes learning in VR special and which challenges learners face, the nature of VRLEs and the resulting learning setting are described in the following section.

2.1.1. Defining virtual reality environments

Virtual Reality (VR) has been defined in many ways throughout the literature due to disciplines emphasizing different characteristics of VR. To date, a standardized approach regarding what can and cannot be considered VR is lacking (Kardong-Edgren, Farra, Alinier, & Young, 2019). For example, VR was defined by Pan and Hamilton (2018) as a *computer-generated world*.

2. Theoretical background and related work

This definition has been criticized as not being specific enough and as not being clearly distinguishable from other concepts, such as augmented reality (Slater, 2018; Kardong-Edgren et al., 2019). Thus, to capture the uniqueness of VR, it must be understood not only as a computer-generated virtual environment, but also as an environment that perceptually surrounds learners to immerse them in the virtual world (Slater, 2018). When the VR environment is accessed through a head-mounted display with the simulation being presented in a stereoscopic way, this is termed *immersive* VR (Kardong-Edgren et al., 2019). This type of VR stands in contrast to low- or non-immersive settings, where learners interact with a 3D desktop simulation (Kardong-Edgren et al., 2019; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). When learning in VR, learners perceive and interact with the learning content in a virtual world. The following definition of VRLEs by Baceviciute, Mottelson, Terkildsen, & Makransky (2020), highlights their differentiated nature:

‘[VR is...]a computer-mediated simulation that is three-dimensional, multisensory, and interactive so that the user’s experience is ‘as if’ inhabiting and acting within an external environment’ (Baceviciute et al., 2020, p. 37).

This definition is employed in the present thesis for several reasons. First, it acknowledges that learners can perceive the VRLE, and consequently, the included learning content, through different sensory channels. It also emphasizes the importance of the active role of the learner and highlights the subjective perception of the learner.

When referring to immersive VR, two basic concepts can be differentiated: *immersion* and *presence* (Mayer, 2020; Makransky & Petersen, 2021). The concept of immersion refers to technological and objective factors of VR. Such factors can include, for example, the technological provision of the learning environment via a certain medium or specific hardware. Depending on the technical setup and the chosen simulation, different degrees of freedom are available. Moreover, depending on the available degrees of freedom, learners can interact with their surroundings and move in the virtual world (Mayer, 2020; Babu, Krishna, Unnikrishnan, & Bhavani, 2018). In contrast, the concept of presence refers to the subjectively experienced state of immersion and is described as the feeling of ‘*being there*’ (Riva, Davide, & IJsselsteijn, 2003, p. 3) during learning in VR (Mayer, 2020). Both concepts, immersion and presence, have the power to substantially impact learning processes (Makransky & Petersen, 2021).

2.1.2. Specific qualities of virtual reality learning environments

The nature of VRLEs offer the opportunity to present learning material in a contemporary and motivational way (Makransky et al., 2020). However, this can also lead to different challenges and potential barriers to learning (Ragan et al., 2015; Mulders et al., 2020; Hu-Au & Okita, 2021; Huang et al., 2020). To investigate the potential of VRLEs, the learning content should be selected in a way that learners can benefit from experiencing it in an immersive manner (Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020). The situated setting, inducing presence, is described as being beneficial for learning outcomes (Makransky, Lilleholt, & Aaby, 2017).

For instance, Thorsteinsson and Page (2008) outlined the fact that learners can constructively perceive the learning environment. By exploring the VRLE, learners might investigate relevant questions and experience learning problems from a new perspective (Y. L. Chen, 2016).

Furthermore, VRLEs can be used to create situated and realistic learning environments for the observation of scientific content (Hedberg & Alexander, 1994). Possible barriers to observing these processes in real laboratories include limited time and financial resources. Some processes take place at the micro-level and therefore cannot be observed in conventional laboratories, as these are invisible to the human eye (Radianti et al., 2020; Miller et al., 2021). For instance, one study of Bell and Fogler (1995) used a VRLE to display chemical reactions in an engineering course. Similarly, in their study, Parong and Mayer (2020) provided insights into the human body and learners could perceive different cells in the blood vessel in an immersive way.

It is also possible to perceive new technology before even building the real prototype and to interact with it in a VRLE (Onyesolu & Eze, 2011). Abstract technical concepts or interactions with robots can be displayed flexibly in VRLEs, allowing learners to interact and familiarize themselves with new technologies in a cost-effective and safety-ensuring way (Hu-Au & Okita, 2021). For instance, learners might observe an operating robot in its specific environment (Vogt, Babel, Hock, Baumann, & Seufert, 2021b, 2021a). Thus, the configuration, moving trajectory, and interaction pattern of a service robot allows learners to perceive the topic of robotics in an immersive way and might thereby enable learners to understand basic concepts more intuitively as they are present in a situated setting (Makransky & Petersen, 2021).

In VRLEs, different sensory channels are used to convey information depending on the chosen presentation format. In general, global distinctions for presenting learning material can be made: First, the codality of the presentation format can be described (Paivio, 1990). This refers to the code used to present the learning content. Here, one can distinguish between verbal code (narrations or printed words) and non-verbal or pictorial code (VR animations or pictures). Second, the modality used to convey information can differ in that information can be presented in a visual or auditory format (Moreno & Mayer, 2007).

Analyzing prior findings, one can conclude that the content, as well as the provided presentation format for learning in VR, need to be chosen carefully (Wu et al., 2020; Makransky & Petersen, 2021). This is consistent with the idea that different forms of presentation such as pictorial or verbal representations are more appropriate for displaying certain contents (Schnotz & Bannert, 2003). Most VRLEs are dominated by visual, pictorial representations that illustrate the learning content. *But what makes pictorial representations particularly attractive?* Pictures are dually coded which results in advantages for storing and retrieving pictorial information (Paivio, 1986). However, previous literature has described that learning materials with animations can overload learners as they must simultaneously keep information active in their working memory (de Koning, Tabbers, Rikers, & Paas, 2010). Hence, animations are described to violate the *apprehension principle*. This principle states that ‘*the structure and content of the external representation should be readily and accurately perceived and comprehend*’ (Tversky, Morrison, & Betrancourt, 2002, p. 255). Animations are often too

2. Theoretical background and related work

complex or too fast to be accurately perceived. Similar effects could occur when learning with VR animation (Makransky & Petersen, 2021). Here, learners would be challenged to process a lot of information at a given time. This overload in combination with little control over the learning environment can lead to suboptimal elaboration processes (Lowe, 2003; Tversky, Heiser, Mackenzie, Lozano, & Morrison, 2008; Tversky et al., 2002). As a result, learners are unable to adequately process the information and thus fail to build a mental model (de Koning, Tabbers, Rikers, & Paas, 2009). Here, the chosen presentation formats displayed in VR might be crucial to avoid overload and to support learners to process the content deeply.

In addition, verbal components such as labels or narratives can also be integrated. Using these different sources of information might trigger the multimedia effect when learning in VR (Mayer, 2014). It is assumed that learning with different presentation formats, such as texts and pictures, leads to higher learning outcomes than learning with a solely textual presentation of the same learning content (Paivio, 1990; Mayer, 2014).

For instance, when presenting new technology such as a new service robot, the VRLE might include realistic animations of the robot, while additional more abstract information might be provided as narration (modality principle; Mayer, 2014). In this setting, learners can observe the functionality of the robot from different perspectives and might see the robot's moving trajectory and how it performs its task (Vogt, Babel, et al., 2021b). Abstract information provided in the narration can be mapped to the displayed information in the VR animation.

In contrast, when learning more abstract scientific content, such as chemical reactions, the VRLE should be designed differently. VRLEs offer the possibility to display both micro as well as macro phenomena (Radianti et al., 2020). This idea of changing flexibly between different levels of learning content is also reflected in the *elaboration theory* developed by Reigeluth and Stein (1983). This theory posits that different approaches can be employed to gain further insights into a given topic: from general-to-detailed. This means that learners must first learn about a global perspective followed by a specific epitome, or in other words, a local aspect of the topic (Reigeluth & Stein, 1983). This switching between micro and macro levels can be easily implemented in a VRLE.

To sum up, VRLEs offer great potential to present learning material in a contemporary way and offer new approaches to learn at universities. However, learners face specific challenges while learning in VR that might impact learning outcomes (Ragan et al., 2015; Mulders et al., 2020; Hu-Au & Okita, 2021; Huang et al., 2020). As a result, many findings related to VRLEs have been published throughout the past years, which are discussed in detail in the following section.

2.2. Empirical findings on learning in virtual reality

Although VR environments are sometimes described as a *new* technology, their history dates back to the 1970's (Mazuryk & Gervautz, 1996). Back then, VR environments were used in the context of flight or military simulations. Older publications analyzing the potential of VR environments in the learning context outlined that the technical implementation of learning

content was not sufficiently developed, which resulted in poor instructional design of VRLEs (Riva et al., 2003; C. J. Chen, Toh, & Ismail, 2005). The review of Di Natale, Repetto, Riva, and Villani (2020) criticized the methodological weaknesses of some studies. For example, such studies often included a lack of randomization, small sample sizes, unbalanced samples, and non-validated measurement instruments. Furthermore, it was postulated that learning in VR was not a promising approach for all topics (Moreno & Mayer, 2004).

Throughout the past few years, the number of publications investigating learning in VR has rapidly increased (Makransky & Petersen, 2021). Taking a closer look at previous findings, many studies compared immersive VRLEs with non-immersive computer-simulated or conventional learning settings, such as learning with a worksheet, PowerPoint slides, videos, or other conventional learning material (Parong & Mayer, 2018; Howard & Lee, 2019; Klingenberg et al., 2020; Webster, 2016). To summarize these studies, Wu et al. (2020) conducted a meta-analysis including 35 studies. They found a small overall beneficial effect of VRLEs compared to other media. However, they outlined that many of the included studies found no or even detrimental effects of learning in immersive VR. Based on these findings, caution regarding the potential of VRLEs as effective learning environments appears warranted.

For instance, Webster (2016) documented that learning in VR resulted in higher learning outcomes compared to learning in a desktop simulation. In contrast, Parong and Mayer (2018) found that learning in VR distracted learners cognitively and thus resulted in a lower learning outcome compared to a non-immersive desktop environment.

A preliminary framework to outline and understand the heterogeneous findings from media comparison studies was thereby developed and is entitled the *Cognitive Affective Model of Immersive Learning* (CAMIL) by Makransky and Petersen (2021). This model will be described in the next subsection.

2.2.1. Cognitive Affective Model of Immersive Learning

The *Cognitive Affective Model of Immersive Learning* (CAMIL) proposed by Makransky and Petersen (2021) presents an overview of important factors for learning in VR (see Figure 2.1). In CAMIL, different influencing factors and their relations to certain learning outcomes are described. As for influencing factors, technological factors, affordances of immersive VR, and cognitive-affective factors are included. Moreover, their impact on learning outcomes is described using specific paths connecting the input to the output factors. On the left side of the model, three technological factors are included: immersion, representational fidelity, and further control factors.

The technological factors impact presence and agency, which constitute the level of controlling actions in VR (Makransky & Petersen, 2021). These two factors represent the affordances of immersive VRLEs. The included affordances have an impact on the six cognitive and affective influencing factors outlined in this model: interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation. As outcome variables, factual knowledge, conceptual knowledge, procedural knowledge, and transfer of learning are included.

In their research, Makransky and Petersen (2021) describe previous findings from media

2. Theoretical background and related work

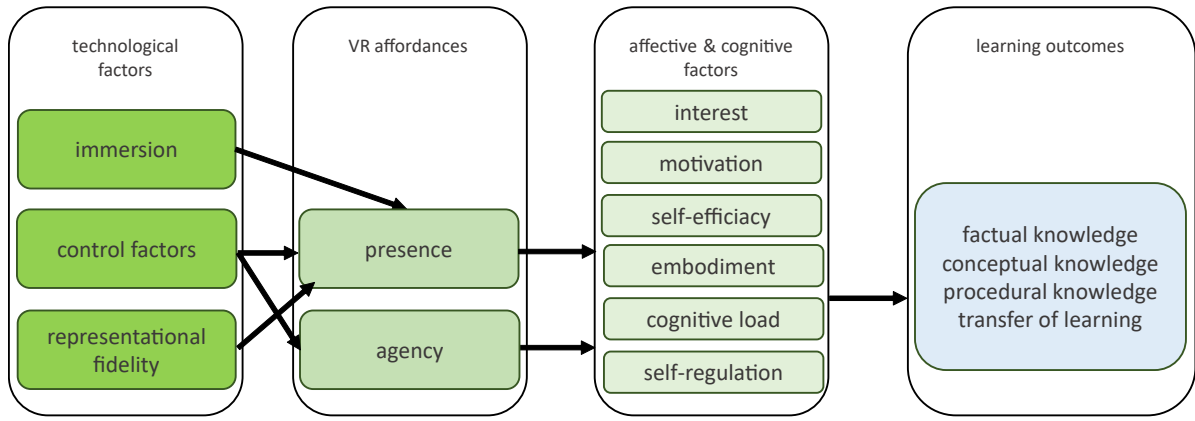


Figure 2.1.: Simplified version of the Cognitive Affective Model of Immersive Learning (Makransky & Peterson, 2021, p. 7)

comparison studies for each included path. For instance, a positive effect of immersion on presence is described (Johnson-Glenberg, 2019). Moreover, higher presence is posited to go along with higher intrinsic motivation, and higher intrinsic motivation is thought to be linked to improved learning outcomes (Harackiewicz, Smith, & Priniski, 2016; Dev, 1997). In contrast, higher presence might also be linked to higher cognitive load, and in turn, hampering effects on learning outcomes might occur (Mayer, 2020).

Additionally, several important assumptions for learning in VR are postulated. For instance, one main assumption of CAMIL is that evidence from research investigating conventional, non-immersive learning environments can be transferred to immersive learning in VR. It is also suggested that media for presenting learning content might interact with how the learning material is presented and potential support approaches.

Although, CAMIL offers important insights into relevant factors for learning in VR, many questions surrounding this topic remain unanswered. For instance, CAMIL does not include any information about the specific underlying cognitive learning processes or how the presentation format might interact with other relevant factors, such as motivation.

The empirical base for developing CAMIL included media comparison studies. When a *new* learning medium gains attention, the first question is always whether it has a universal advantage over traditional learning media (R. E. Clark, 1983). This is also reflected in the research tradition on VR in recent years. However, media comparison studies are typically viewed in a critical light, especially due to central problems regarding the guaranteed comparability of different media (R. E. Clark, 1983). This critical view on findings of media comparison studies in the context of VRLEs simultaneously highlights the weaknesses of media comparison studies.

For instance, Parong and Mayer (2020) used a VRLE that contained continuous VR-animations in combination with spoken narration, whereas the presented non-immersive desktop environment simply displayed PowerPoint slides with written text. Thus, this study incorporates a major problem of media comparison studies: different characteristics of the used media might bias the overall conclusion that one media is more efficient than the other (R. E. Clark, 1983). Furthermore, media as well as instructional settings including different approaches to

supporting learners might impact learning outcomes, and interactions between these factors are possible (Parong & Mayer, 2020; Makransky & Petersen, 2021).

Taken together, to explore the potential of VRLEs, the focus should be less on media comparisons and more on the medium of VR itself. Central questions to this end include: *What are the specific challenges that may stem from this kind of learning environment? How can learners be supported to successfully master these challenges?* These questions will be discussed in the following sections.

2.2.2. Challenges while learning in virtual reality

One important characteristic of VRLEs is the size of the visual field where learning content can be displayed. In conventional settings, space where learning content is presented is more obvious and limited to PowerPoint slides, the book or computer being used (Makransky et al., 2020). Consequently, the size of the visual field of view might challenge learners to select relevant information (Ragan et al., 2015; Mulders et al., 2020). Through this, additional resources might be consumed in order to locate relevant information (Makransky, Terkildsen, & Mayer, 2019). When VRLEs display the learning content in such a way that it switches between micro and macro levels, this might additionally pose challenges for learners regarding how they should guide their attention. In this case, learners must once again select the relevant information, or its presentation in VR, as the virtual world changes perspective. In specific terms, this means that each time they zoom in or out, they must reorient themselves, which may involve navigating through the VRLE (Vogt, Albus, & Seufert, 2021). Furthermore, as a realistic VRLE might contain elements that are necessary to present the learning content in a situated way, but do not convey learning content, these might distract learners from processing the learning content (Howard & Lee, 2019). In a conventional setting it is more obvious for learners where the learning content is presented and what the surrounding elements are (Howard & Lee, 2019). Hence, learners are particularly challenged to guide their attention to important elements in VRLEs (Albus, Vogt, & Seufert, 2021).

Moreover, when VRLEs are implemented in a constructivist way where learners can explore the virtual world, guiding attention might be challenging and learners might explore the environment in a superficial manner (Mayer, 2004). As a result, VRLEs can reduce learner's focus on the learning content and the resources provided to deeply process the learning content (Howard & Lee, 2019). One study published by Mayer (2004) showed that learning with simulations can have harmful effects when the behavioral or exploratory activity replaces learner's efforts to engage deeply with the learning content. Hence, learning behavior can be extended by additional qualities, and pre-existing learning behavior from conventional settings need to be specifically adapted to the learning settings in VRLEs.

Once learners have mastered the challenge of locating the relevant information, additional challenges in the learning process may follow. Previous research has found that VRLEs often use a combination of static images or VR animation and verbal elements to convey knowledge. However, using different presentation formats might pose challenges for the learners in terms of being able to connect information between these different formats (Vogt, Klepsch, Baetge, &

2. Theoretical background and related work

Seufert, 2020; D. Gentner, 1983).

Learners must first process the individual presentation formats and understand their content before they can recognize potential content overlaps between individual formats. This is not only a matter of recognizing certain elements and perceiving them as belonging together (element-to-element mapping), but can also involve parallels at the process level (relation-to-relation mapping), which require a deeper understanding of the learning content (Díaz & Ioannou, 2019; D. Gentner, 1983). This can be particularly challenging for learners in VRLEs that are not self-paced. Thus, to overcome superficial processing of the learning content presented in VR, learners might need to be supported.

2.2.3. Fostering deep learning in virtual reality

Presenting learning content in VR does guarantee deep processing of the provided content (Vogt, Babel, et al., 2021b; Wu et al., 2020). Surface level processing while learning with interactive technology can be contrasted with deep level processing. Díaz and Ioannou (2019) describe superficial processing as simple recall or memorization without further engaging into the learning processes. In contrast, deep level processing of the learning content also involves the deduction of meaning of the presented content, self-monitoring, and a commitment to learning (Biggs, 1989). This commitment involves gaining a deeper understanding and critically reflecting on the learning content. In the context of learning, Díaz and Ioannou (2019) characterize the term *deep learning* as follows:

‘Deep learning is also about developing habits to consistently think and reflect, approach new phenomena in thoughtful ways, and see new phenomena from different perspectives in everyday life’ (Díaz & Ioannou, 2019, p. 16).

According to this definition, learning content must be processed semantically in order to categorize the learning process as deep learning¹.

To overcome the challenge of surface level processing while learning in VR, previous studies have investigated different support approaches. As a basic distinction of these support approaches, one can distinguish between external support approaches and internal support approaches. External support approaches refer to the specific design of the learning environment and the presentation formats used to convey the learning content (Jenkins, 1979). Internal support approaches incorporate certain elements into the learning environment that stimulate the active role of the learner and trigger learners to show appropriate learning behavior (Jenkins, 1979).

One external support approach constitutes adding information directly related to the learning content prior to the start of the learning unit. This approach is called pre-training (Mayer, 2017). By adding this information, learners can reactivate prior knowledge and have a brief

¹The term *deep learning* is used by different disciplines. For example, in computer science, the term deep learning refers to a special method of information processing and a subfield of machine learning. In machine learning, deep learning refers to deep neural networks that are used to analyze large data sets. The present thesis refers to deep learning in the context of educational psychology and therefore does not deal with deep learning in the context of neural networks.

preview of *what to learn* (Vogt, Babel, et al., 2021b; Meyer et al., 2019). For instance, the study of Meyer et al. (2019) found a beneficial effect of pre-training on different levels of the learning outcome, indicating that pre-training might help learners to cope with the requirements of the VRLE. It should, however, be noted that within this study, the VR learners were compared to learners with videos. When learning with videos, the beneficial effect of pre-training was not found (Meyer et al., 2019).

Another global approach to support learners in VR is to use internal support approaches that trigger learners to adequately adapt their learning behavior. For instance, Klingenberg et al. (2020) added elements in a VRLE to trigger generative learning strategies. They found beneficial effects of this support approach on learning outcomes. Furthermore, Parong and Mayer (2020) used summary prompts after each learning session in VR and found such prompts to support learning in VR.

From these findings, it can be concluded that promising, preliminary approaches exist to support learners in VR. The approaches described thus far suggest that there is a need for further research in this area. For example, the use of pre-training is a rather content-bound approach. Here, the domain-specific prior knowledge that can help learners to better understand the learning content in VR must be individually evaluated. Moreover, when using summary prompts, learners are required to interrupt their learning unit to take notes. Thus, the question arises of whether other promising approaches might exist that do not contain such problems.

To answer this question, the underlying learning processes must be comprehensively understood. For instance, CAMIL describes certain influencing factors on learning outcomes and thus describes the impact of input on output variables. However, it does not focus on the underlying cognitive learning processes and thus does not deliver insights into the consequences of different support approaches on the cognitive subprocesses while learning. Furthermore, it does not become apparent whether these support approaches directly address the specific challenges the learners face while learning in VR. In the following section, methods through which these underlying cognitive processes and potential influencing factors can be investigated in a sophisticated and rigorous manner are discussed.

2.3. Relevant factors for fostering learning in virtual reality learning environments

To develop a better understanding of how to overcome challenges while learning in VR by support approaches, the related underlying cognitive learning processes require further research. As recommended by Jenkins (1979) four factors should be considered when conducting memory experiments to investigate processes such as learning. These factors are translated into the learning context and can be seen in Figure 2.2.

Jenkins (1979) explained that experimental findings are highly context-sensitive. Accordingly, these influencing factors should either be kept as constant as possible or systematically manipulated (Roediger, 2008).

The first factor described by Jenkins (1979) constitutes the chosen learning setting (F1). To

2. Theoretical background and related work

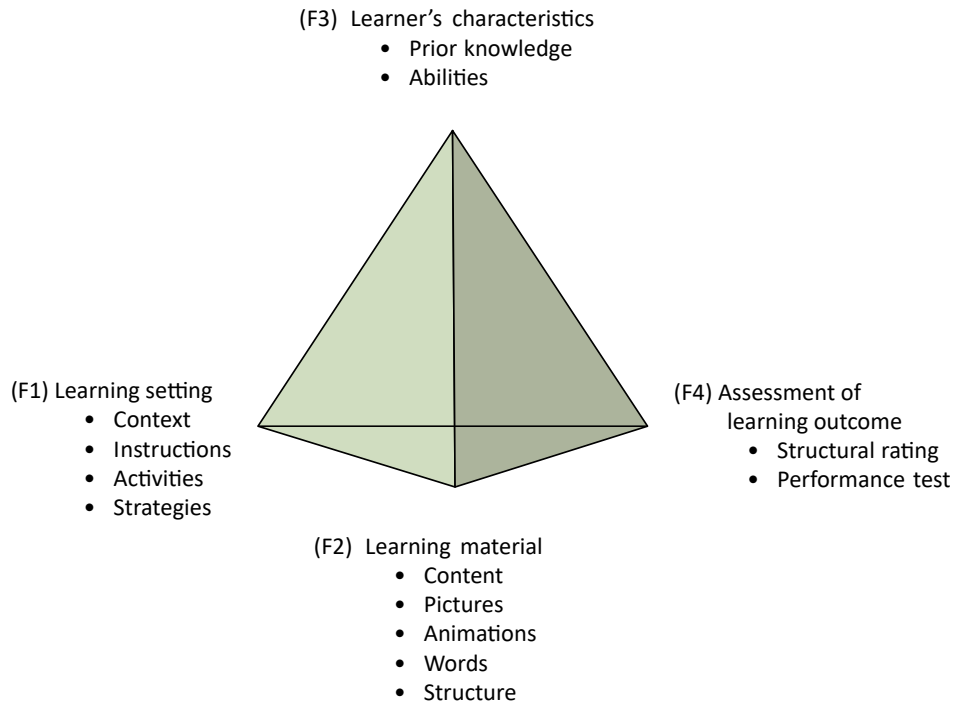


Figure 2.2.: Adapted version of the Tetrahedral Model of Memory Experiments by Jenkins (1979)

this end, the underlying cognitive learning processes are thought to depend on the specific context and the provided instructions. Additionally, the scope of actions available for learners might impact the learning processes, as learners show different behavior depending on the provided boundary conditions. Furthermore, they might apply suitable strategies that depend on the learning setting. Both, learning behavior and learning strategies can be described as internal factors that impact learning processes. This factor can be classified as an internal support approach. In the context of the present thesis, immersive VRLEs were chosen as learning settings.

The second factor focuses on the provided learning material (F2) in more detail. The chosen presentation formats for the learning content, for example, combining text and pictures or VR animations, as well as sequencing these formats, are posited to impact the underlying cognitive learning processes. This factor can be characterized as an external factor influencing learning processes and thus external support approaches are related to this factor.

As the design and the chosen presentation formats of the VRLE have an impact on the underlying cognitive learning processes and the resources that learners need to invest to process the given learning content, the design needs to be chosen carefully. One important issue that should be considered is that the presentation format should match the chosen content, and through this, increase *computational efficiency* (Larkin & Simon, 1987; Hegarty, 2011). The concept of computational efficacy describes the fact that two different presentation formats containing the same learning content do not involve the same expenditure of cognitive resources to process the relevant information (Larkin & Simon, 1987). Thus, depending on the content, it may be more efficient to choose a visual or textual presentation format (Johnson-Laird, 1983).

Several key factors influencing computational efficiency exist: the structure of the learning content presented, search and recognition patterns, and integration of new information into the previous knowledge network (Larkin & Simon, 1987).

The third relevant factor presented by Jenkins (1979) represents the critical learner's characteristics (F3). The importance of this factor for learning has been outlined by many researchers (Park, Münzer, Seufert, & Brünken, 2016). As the underlying cognitive learning processes do not only depend on the chosen setting and learning material, but also on characteristics of the learners, such characteristics should also be examined in this context. This is especially true as whether different presentation formats foster deep learning processes depends on learner's characteristics. In the context of learning, prior knowledge is described to have a profound impact on underlying cognitive learning processes as well as on resulting learning outcomes (Vogt et al., 2020; Seufert, 2019). Furthermore, verbal or spatial abilities might explain a substantial amount of variance in learning performance (Lee & Wong, 2014; Schipolowski, Wilhelm, & Schroeders, 2014).

The fourth factor described by Jenkins (1979) refers to how to measure learning outcomes (F4). To this end, several approaches can be used to measure learning outcomes. Depending on the chosen measurement approach, different effects of support approaches can be expected. Additionally, the chosen measurement impacts the extent to which one gains insights into underlying cognitive learning process. When measuring learning outcomes, one might focus on performance measurements or on the configuration of the resulting mental model (Bloom, 1956; Trumpower & Vanapalli, 2016). In line with this idea, the fourth factor recommended by Jenkins (1979) is the chosen assessment of the learning outcome.

To summarize, relevant factors that should be taken into account when examining cognitive processes have now been outlined. The central element in Jenkins' (1979) *Problem Pyramid*, the underlying cognitive learning processes themselves, will now be deepened in the next section.

2.4. Underlying cognitive learning processes

Based on the work of Makransky and Petersen (2021), it is assumed that basic learning processes investigated by use of conventional learning materials can be transferred to the context of learning in VR. In connection to this, important aspects of pre-existing models and theories describing cognitive learning processes are outlined in the following section. Based on these cognitive learning processes, the specific challenges of learning in VRLEs can be characterized in more detail, and promising support approaches can be deduced.

2.4.1. Cognitive Theory of Multimedia Learning

To describe such underlying cognitive processes, the *Cognitive Theory of Multimedia Learning* (CTML) proposed by Mayer (2005) is often used. This theory describes the perception, selection, organization, and integration of learning content in one mental model (see Figure 2.3).

The model assumes three different memory systems: sensory, working, and long-term memory. Two different channels are available, one for processing pictures, and one for processing words.

2. Theoretical background and related work

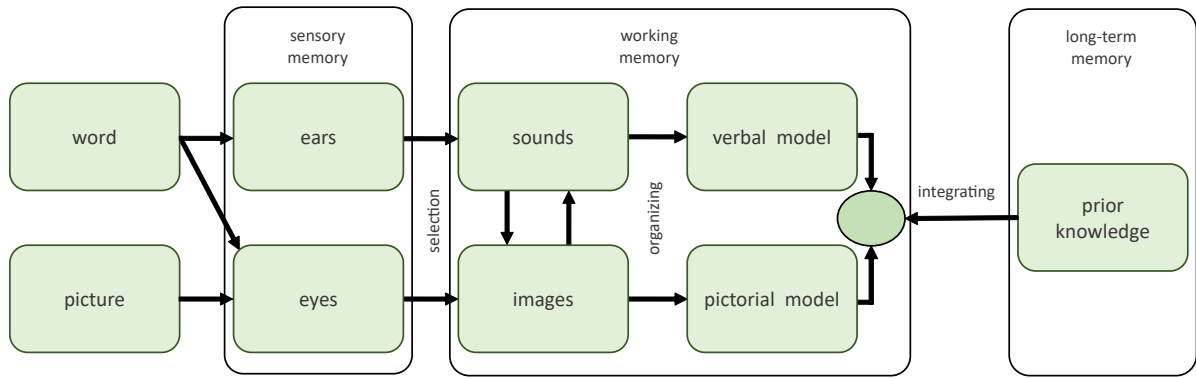


Figure 2.3.: Simplified version of the Cognitive Theory of Multimedia Learning (Mayer, 2020, p. 40)

When pictures are processed, they are dually coded (Paivio, 1990). This is based on the fact that pictures are automatically translated into verbal code and are therefore processed in both channels, resulting in advantages regarding the storage and retrieval of pictorial information.

After perceiving information via eyes or ears, the information is then temporarily stored in the sensory memory. Before the information is further processed and organized in the working memory, only the relevant information is selected as the capacity of the working memory is limited (Baddeley, 1992). This limitation is a decisive factor for the extent to which a coherent mental model can result: Learners need to have sufficient resources to process all relevant information. From processing and organizing the learning content two separate models are constructed: one for pictures and one for verbal information. In the end, both verbal and pictorial models are integrated into one mental model under consideration of prior knowledge. Thus, the CTML describes learning as an active process where learners need to engage and not simply passively perceive the learning content (Mayer, 2005).

Although the CTML included some details related to cognitive learning processes, important factors were not considered. As a result, the CTML was extended by (Moreno & Mayer, 2007) to the *Cognitive-Affective Theory of Learning with Media* (CATLM), which adds further influencing factors such as motivation and self-regulation. These factors might have an impact on selecting and organizing information while learning, and may also influence the integration of new information into the mental model. Hence, affective or motivational factors might moderate or mediate the underlying cognitive learning processes (Moreno & Mayer, 2007; Leutner, 2014).

2.4.2. Integrated Model of Text and Picture Comprehension

In contrast to the CTML, the *Integrated Model of Text and Picture Comprehension* (IMTPC) postulated by Schnotz and Bannert (2003) incorporates more information about the integration process for textual or verbal and pictorial information. In this model, the different processes that are necessary for the cognitive processing of textual and pictorial representations are described in detail. For the construction of a mental model, the steps of selection, organization, mapping of analogous structures, and, if necessary, the parsing of symbolic structures are

described (see Figure 2.4).

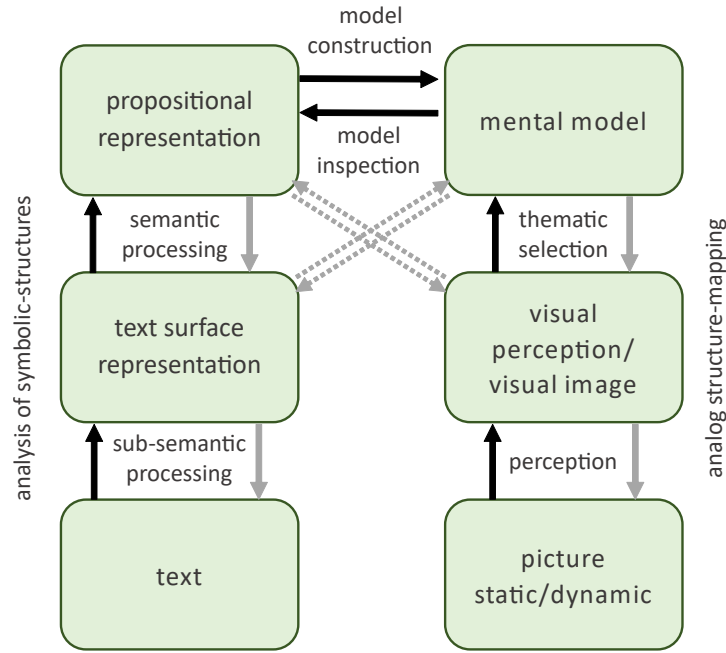


Figure 2.4.: Simplified version of the Integrated Model of Text and Picture Comprehension (Schnotz & Bannert, 2003, p. 145)

This model assumes separate branches for textual and pictorial information. Information is processed by two basic mechanisms: task-oriented top-down processing that is highly dependent on pre-existing cognitive schemata (Ullmann, 1984), and bottom-up processing that is closely related to the presentation of the information. Both processes have the function of selecting and organizing information. Textual and pictorial representations must first be processed superficially or sub-semantically. This results in an internal text surface representation or an internal visual image (Schnotz & Bannert, 2003). After this, when enough cognitive resources are available, semantic processing of the internal representations takes place (Baddeley, 1992; Mayer, Mathias, & Wetzell, 2002). Here, the IMTPC outlines the importance of the code of information. Pictorial information is *analog* and can therefore be directly integrated after thematic selection into its analog counterpart, the mental model. In contrast, textual information, which is classified as *symbolic*, needs to be translated or parsed before being integrated into the mental model (Schnotz & Bannert, 2003). The semantic processing of the internal text surface results in a propositional representation. This propositional representation no longer contains the exact syntax of the former external representation, but rather includes premises with a certain predicate-argument form (Johnson-Laird, 1981). To include this information into the mental model, parsing the symbolic code is necessary. Hence, an additional effortful process is required to include textual information into the mental model. Based on cognitive schemata, a continuous process of model construction and model inspection is assumed (Schnotz & Bannert, 2003). Thus, both, the mental model and the propositional representation influence each other. The IMTPC thereby disagrees with the assumed one-to-one mapping process for the integration of textual and pictorial described in the CTML (Mayer, 2005; Schnotz & Bannert, 2003). Based on

2. Theoretical background and related work

the processes described, one could assume that using pictorial representation would be superior in general. However, this is not the case: For instance, pictures are suitable for representing inferences, while texts are suitable for representing abstract details (Johnson-Laird, 1981). As previously outlined, this aspect of the choice of learning material is also discussed concerning the concept of computational efficiency (Simon, 1978). Assuming that a pictorial and a textual representation are equivalent from an informational perspective, the effort to process the given information more deeply can differ significantly (Hegarty, 2011). This depends not only on the used analog or symbolic code of the representation, but also on how well certain patterns or inferences can be recognized, or to what extent control strategies for the search for relevant information are needed (Larkin & Simon, 1987). Thus, when two representations display the same content in a way that can be accessed with the same effort and at the same time, the computational efficiency for both representations is assumed to be equivalent (Larkin & Simon, 1987; Simon, 1978).

Another important aspect of the IMTPC involves potential shortcuts: sub-semantically processed textual learning content might be translated and integrated directly into the mental model without semantically processing this information and building a propositional representation (Schnotz & Bannert, 2003). As a consequence, certain concepts might be classified as closely related based on the superficially processed syntax, but learners would not be able to deduce correct cause-and-effect chains between the different concepts (Vogt, Babel, et al., 2021b).

After providing detailed insight into the underlying cognitive learning processes, we can go back to the question of how to support learners in VRLEs and can have a look at the different factors proposed by Jenkins (1979) that might impact the underlying processes.

2.5. Designing appropriate learning material

Based on Jenkins (1979), the specific design of the chosen learning material (F2) might substantially affect the underlying cognitive learning processes. Depending on the chosen presentation format, different design approaches are available, and learners might be supported by different design features. The effect of the specific design and the chosen presentation format has been classified as an external support approach to ease learning in VR.

VRLEs can be visually complex, as they display an extended field of view (Ragan et al., 2015; Mulders et al., 2020). Thus, learners might need guidance to process the relevant information (Ozcelik, Arslan-Ari, & Cagiltay, 2010). This is particularly the case when the VRLE switches between micro and macro levels. For instance, by using guiding elements such as *signals*, the probability of processing relevant information is increased (Mason, Pluchino, & Tornatora, 2013). Consequently, learners might have more time to focus on the relevant information. Furthermore, learners would theoretically have more cognitive resources left, as unnecessary search processes would be reduced (de Koning, Tabbers, Rikers, & Paas, 2007).

After having selected the relevant information, this information must be organized and again prioritized (Mayer, 2014). When only the relevant information is selected, this might also ease

the process of organizing new information, as this should be more consistent and coherent when only task-relevant information is included (Schnotz & Bannert, 2003). Again, lower cognitive strain while processing the learning content might emerge when learners are supported in the selection processes in the consecutive process of organizing learning content (R. C. Clark, Nguyen, & Sweller, 2011).

To foster the integration of information, different support approaches can be selected (Vogt, Albus, & Seufert, 2021; Vogt, Babel, et al., 2021a, 2021b). It may prove beneficial to provide learners with help on the surface to retrieve relevant information. Consequently, learners could also experience beneficial effects on their subsequent cognitive processing, as only relevant information is organized and integrated (Mayer, 2014). On the contrary, it may be necessary to support learners with integrating corresponding elements into the mental model, especially if they stem from different presentation formats (D. Gentner, 1983; van der Meij & de Jong, 2006). Depending on the support approach used to select relevant information, finding corresponding information in the different representations might become easier, and thereby support both mapping corresponding elements and relations (D. Gentner, 1983). Through this, the inclusion of new information into the pre-existing knowledge network might be eased (Jonassen & Henning, 1999). Before adding supportive elements into the VRLE, however, specific characteristics should be considered and potential challenges defined in order to choose a support approach that also helps with the critical underlying cognitive learning processes.

2.5.1. Adding annotations in virtual reality

Using a VRLE with zooming in and out on micro and macro level might create visual complexity that challenges learners. Thus, adding highlights or signals to guide the learners' visual attention is a promising approach to foster learning processes (Mautone & Mayer, 2001; de Koning et al., 2007). These signals incorporate two different functions: reducing unnecessary visual search and guiding attention to relevant content (de Koning et al., 2007; Ozcelik et al., 2010). Previous studies using multimedia learning environments have outlined the beneficial effects of signaling on learning outcomes (van Gog, 2014; Alpizar, Adesope, & Wong, 2020). For instance, medium beneficial effects with the size of $d = 0.31$ were reported for signals using color contrast (Alpizar et al., 2020). Heterogenous effect sizes were reported to depend on different moderators such as prior knowledge or level of signaling. However, when simply guiding attention, learners are still challenged to connect certain components from micro and macro levels or to map information from narration and VR animation. As a result, learners might not be able to recognize and map the presented concepts in the learning environment. Extending simple visual signals to labels or annotations that guide attention and help to map different representations in the learning material may thereby be even more effective (McTigue, 2009; Vogt, Albus, & Seufert, 2021). Previous studies described a positive effect of annotation on learning outcomes when inserted in pictorial representations (McTigue, 2009). Annotations fostered learning scientific concepts: Learners with annotated pictures showed higher knowledge at the comprehension level than learners who learned with non-annotated pictures. Moreover, Mayer and Gallini (1990) found positive effects of annotated pictures in learning processes

2. Theoretical background and related work

and substeps. Therefore, including textual annotations into the VRLEs might extend the benefits of using simple superficial signals to guide attention. For instance, when the pictorial VR animation is combined with an auditive narration, learners might assign the annotated pictorial representation with the terms described in the narration (Mayer, Steinhoff, Bower, & Mars, 1995). This enables learners to map corresponding elements in the learning content between different representations (D. Gentner, 1983) and may facilitate the integration of the new information of the different representational formats into the mental model. Additionally, annotations repeat the terms of the narration and might therefore foster learning of basic definitions (Vogt, Albus, & Seufert, 2021). In particular, learning outcomes on the recall level might be increased when using annotations, as simple repetition is most effective for learning facts (Boers, Warren, Grimshaw, & Siyanova-Chanturia, 2017).

To sum up, annotations can be expected to reduce visual search, guide attention, help to integrate and map information, and support simple recall due to redundant terms in the learning environments (Vogt, Albus, & Seufert, 2021). Specifically, in visually and perceptually demanding VRLEs that display abstract concepts, these annotations might be a valid approach to support learners to process and integrate learning content and to build a coherent mental model.

2.5.2. Sequencing textual and pictorial representations

Research suggests that the simultaneous presentation of pictorial and textual formats is beneficial for learners (multimedia principle; Mayer, 2014)– particularly when related concepts are presented closely together, both spatially and temporally (continuity principle; Mayer, 2014). Then, both textual and pictorial information is active at the same time in the working memory. However, presenting both representations simultaneously is only beneficial when pictorial and textual information can be mapped. In other words: the information included in the different representations needs to include corresponding or complementary information (Schnotz & Bannert, 2003). Although presenting pictorial and textual information might be beneficial when considering certain boundary conditions, this might also go along with cognitive, visual, or perceptual overload (Makransky & Petersen, 2021). As learners must process all information simultaneously, this might result in superficial processing (Baddeley, 1992; Mayer et al., 2002). One method to avoid a potential overload is to sequence the learning material (Eitel & Scheiter, 2014). Furthermore, pictorial presentations might contain different information that cannot be presented with a textual representation and vice versa. For instance, pictures or VR animations can serve as structuring elements for the construction of a mental model (Ainsworth, 2006; Schnotz & Bannert, 2003). Previous findings indicate that when presenting pictorial information first, learners experience increased fixation time on such representations (Mason et al., 2013). Thus, it can be assumed that the information presented in the picture was more deeply processed (Mason, Pluchino, & Tornatora, 2016; Mayer, 2010). Specifically, presenting pictorial information was linked with a positive effect on the integration of the learning content, as well as a higher learning outcome on the transfer level (Mason et al., 2013). This effect was additionally strengthened when learners were previously shown an expert

example (Eye Movement Modeling Examples) of image viewing (Mason et al., 2013; van Gog, 2014). When presenting pictorial representations before textual representations, learners can build a scaffold of the topic and add further details in the learning process. However, only 19 of 42 studies reviewed by Eitel and Scheiter (2014) described beneficial effects of presenting pictorial representations before textual representations. Eitel and Scheiter (2014) describe further characteristics of the learning material that can impact whether presenting pictorial representations may be followed by textual representations. Based on findings of Arndt, Schüler, and Scheiter (2019), additional consideration of the specificity of the used learning content in the respective representation is needed. They additionally explained that presenting a brief pictorial overview, followed by a more specific textual representation can increase learning outcomes (Arndt et al., 2019).

The aforementioned logic can be transferred to the content of VR using the example of exploring the learning content in the virtual world with VR animation. Within this, learners might be provided with a brief overview of the topic and a scaffold of certain components. After this, learners could add further information based on textual information such as a narration without being in the virtual world. Particularly, when presenting learning content that includes new technology, learning can be eased by first providing a brief overview. For instance, displaying a new service robot in the VRLE might allow learners to observe its functionality on a more general level where they might observe different functions of their moving trajectory. Afterwards, more details about the underlying control circuits and related concepts could be provided in a narration. When textual details are described prior to learners having seen the robot in a VR animation, this might consume additional resources, as they might build an analog representation while listening to the text. When presenting pictorial information first, learners might build a brief scaffold based on the VR animation and fill in further details based on the narration.

However, while designing the learning environment appropriately and supporting learners by adding annotations or adequate sequencing might ease learning processes and increase computational efficiency, this does not guarantee learning success. As outlined in the CTML, the learners need to be actively participating while learning, and affective or motivational factors might moderate the underlying cognitive learning processes (Moreno, 2006). This is also reflected in the factor learning setting (F1; Jenkins, 1979). Following this line of reason, the active role of the learners and potential consequences on learning outcomes will be discussed in the next section.

2.6. Active role of learners

Learners need to actively engage in order to select, organize, and integrate the learning content. Fiorella and Mayer (2015) describe different generative learning activities that learners might foster to process learning material. These activities and instructions are also reflected in the factor learning setting (F1) of Jenkins (1979). Many different approaches to stimulate learners to actively engage in learning exist. For instance, learners might make sense of the learning

2. Theoretical background and related work

content by comparing the new information with their prior knowledge or self-explanation of the content (Bisra, Liu, Nesbit, Salimi, & Winne, 2018). Here, one challenge constitutes adapting the different learning activities to the respective learning context (Fiorella & Mayer, 2015). However, exhibiting appropriate learning behavior can be challenging for learners. Hence, strategic activators can be implemented into the learning environment to foster adequate learning behavior and, in turn, to increase learning outcomes (Azevedo, Taub, & Mudrick, 2017).

2.6.1. Activating learners by cognitive prompts

Fiorella and Mayer (2015) describe different learning activities that can be useful for fostering learning and deep processing of learning content. These so-called *generative learning strategies* might be triggered by certain activators in the learning session. One approach to help trigger generative learning strategies and thus adequate learning behavior is to use *prompts* (Kamarainen et al., 2018). Prompts can be characterized as instructional retrieval and/or performance aids (Bannert & Reimann, 2011). Prompts indicate to learners which learning strategies should be used and/or when the use of a specific strategy will result in efficient support of the learning process (Berthold, Nückles, & Renkl, 2007; Thillmann, Kunsting, Wirth, & Leutner, 2009). Thus, prompts can also be described as strategic activators (N. Gentner & Seufert, 2020; Reigeluth & Stein, 1983). As for most learners, VRLEs constitute a new learning setting, they might additionally be challenged to show adequate learning behavior in these environments (Hu-Au & Okita, 2021; Makransky et al., 2020; Huang et al., 2020). Furthermore, learning environments that offer higher degrees of freedom induce more autonomy where learners are required to regulate themselves to choose learning-related activities (Makransky et al., 2020). Depending on the requirements of the learning setting, prompts might stimulate very different learning-related processes (Schumacher & Ifenthaler, 2021).

For example, deeper processing of the learning content can be stimulated to avoid superficial processing of the learning content. These kinds of prompts are also called elaboration prompts, e.g., ‘*Please refer examples from your own life, which illustrate the learning material, are consistent with it, or stand in conflict with it*’ (Endres, Carpenter, Martin, & Renkl, 2017, p. 15). In addition, there are other levels on which the learner can be supported. For example, the learning process can be monitored and reflected by metacognition, e.g., ‘*Think about it: Is your current learning effective?*’ (Daumiller & Dresel, 2019, p. 43). Thus, individual metacognitive strategies can be selected to improve the overall learning performance. Furthermore, strategies on the motivational level can also be applied, e.g., ‘*How great is your motivation for the current learning and how could you possibly increase it?*’ (Daumiller & Dresel, 2019, p. 44). Particularly, when learners do not spontaneously show adequate learning behavior, prompts might foster the use of efficient strategies and thereby increase learning outcomes (Bannert & Reimann, 2011). Additionally, the use of different types of prompts might result in synergic effects (Wischgoll, 2016). Related to the question of *what to prompt* is the question of *when to prompt* (Thillmann et al., 2009). Previous findings describe the beneficial effects of prompting directly integrated into the learning session. However, an interaction between time of prompting and strategy use

was also reported (Thillmann et al., 2009). Consequently, the optimal time to prompt is also context-sensitive and might cause additional load (de Bruin, Roelle, Carpenter, & Baars, 2020).

One major challenge for learning in VR is to overcome superficial processing and exploring the technical setting provided in the VRLEs (Hu-Au & Okita, 2021). Using prompts to stimulate the elaboration of the learning content is thus a particularly promising approach to increase learning outcomes in this context. However, presenting the prompts while learners are in the VRLE might impose additional load and might interfere with processing the learning content (Klepsch, Schmitz, & Seufert, 2017). Additionally, including prompts might reduce the presence and this might reduce learning outcomes (Makransky et al., 2017). In line with the pre-training principle (Mayer, 2014), beneficial effects might also occur because learners have access to relevant information beforehand. In the case of pre-training, this relevant information contains content that is directly related to the learning content (Meyer et al., 2019). In contrast, the prompt does not include learning content. In VRLEs, learners might lack efficient strategies for learning in VR (Huang et al., 2020). Including an elaboration prompt into the instruction before learning in VR might therefore be particularly beneficial as learners will then be prepared to show the respective learning behavior when they enter the VRLE (Vogt, Babel, et al., 2021b). This elaboration prompt might include making sense of the given learning content by thinking of everyday examples related to the topic (Berthold et al., 2007). Furthermore, searching for connections between the presented facts and concepts might foster the integration of the information into a complex knowledge network.

Another approach is to use self- or peer-explanation (Bisra et al., 2018). Within this, learners are required to process the learning content in a way that facilitates a coherent understanding of the topic. This includes, for example, translating the content into their own words (Díaz & Ioannou, 2019). Learners can then differentiate different aspects, connect concepts, and transfer the learning content to other contexts. Prompts are described as being especially effective when including such explanatory components (Berthold, Eysink, & Renkl, 2008). The incorporation of these strategies in an elaboration prompt might therefore stimulate deep learning in VR, as these might activate learners to engage in learning (Vogt, Babel, et al., 2021b). However, whether or not these support approaches are beneficial also depends on the learners and their characteristics, as some learners are already highly motivated and thus spontaneously show effective learning behavior (Vogt, Albus, & Seufert, 2021).

So far, two of the four factors of Jenkins (1979) and their consequences on the underlying cognitive learning processes have been described. One important factor that might also impact the impact of the chosen design and the activation of learners on learning outcomes is the learners themselves (F3). This important factor is also included in the model of Jenkins (1979) and will be outlined in the following section.

2.7. Influence of learner's characteristics

As one of the four important factors for conducting memory experiments, the consideration of learner's characteristics (F3) is mentioned by Jenkins (1979). Thus, to analyze the potential

2. Theoretical background and related work

beneficial effects of support approaches, learner’s characteristics must also be considered. There are several characteristics that differ substantially between learners (Dörrenbächer & Perels, 2016; Seufert, 2019). All models and theories presented thus far in the present thesis (i.e., the CATML, the CTML, the CAMIL, and the IMTPC), included certain aspects that hint at the importance of learner’s characteristics and their relation to the underlying cognitive learning processes. The CATML assumes that motivation might impact all steps of processing the learning content from a cognitive perspective (Moreno & Mayer, 2007). The CTML as and the IMTPC describe prior knowledge to influence underlying cognitive learning processes (Mayer, 2005; Schnotz & Bannert, 2003). Potential aptitude-treatment interactions are well-established and have already been described by Snow (1989). Here, occurring interactions between the chosen instructional design and learner’s aptitudes are considered to cause different learning outcomes. The aptitude and the treatment factor might interact in different ways: Aptitudes might function as an *enhancer* but might also as a *compensator* (Mayer & Sims, 1994). Additionally, certain support approaches might interfere with the given aptitudes of learners (Kalyuga, 2009), and the active role of the learners might be related to their characteristics (Taub, Azevedo, Bouchet, & Khosravifar, 2014). When learners are interested in a certain topic, they might already be activated and intrinsically motivated, and thereby may not benefit from an additional activation or guidance (Vogt, Albus, & Seufert, 2021). Based on the presented theories and models, as well as previous findings, two important learner characteristics are focused on in the present thesis: intrinsic motivation and prior knowledge.

2.7.1. Intrinsic motivation

Intrinsic motivation is described as

‘the desire or intention to learn certain contents or skills’ (Schiefele, 1996, p. 50, translated)

In the learning context, this would mean, for example, that the learner deals with the topic from their own initiative (Rheinberg, 2006). This concept is contrasted with extrinsic motivation, which describes taking action only because a certain positive effect or external reward is expected (Krapp, 1999).

A learner with high intrinsic motivation tends to persist longer on learning tasks and to engage more in learning as a consequence of having more available resources (Ryan & Deci, 2000a; Lepper, Corpus, & Iyengar, 2005). Motivated learners show better self-regulation while learning (Ryan & Deci, 2000b), tend select more appropriate learning behavior, and show more efficient visual search patterns (Miranda & Palmer, 2013). Hence, intrinsic motivation is an important success factor for learning in VR.

Makransky and Petersen (2021) described that intrinsic motivation might be triggered by high fidelity and the level of presence within the VLRE, and may thereby lead to increased learning performance. While some learners might be challenged by the high perceptual load of this setting, intrinsically motivated learners have the necessary resources to succeed. Hence,

learners with low intrinsic motivation might benefit from additional guidance, while this may be less necessarily for learners who already have high intrinsic motivation (Vogt, Albus, & Seufert, 2021). For instance, learners with high intrinsic motivation might have a more efficient visual search, and intrinsic motivation might therefore determine how successful the learning process is (Miranda & Palmer, 2013). Additionally, intrinsically motivated learners can provide more cognitive resources for organizing and integrating information and thus, higher learning outcomes can be expected (Díaz & Ioannou, 2019). Taken together, moderation effects of intrinsic motivation are theoretically plausible and should be considered when adding annotations in learning material (Vogt, Albus, & Seufert, 2021).

2.7.2. Prior knowledge

In the CTML and IMTPC, the importance and potential benefits of prior knowledge for developing a mental model are described (Mayer, 2005; Schnotz & Bannert, 2003). However, prior knowledge can have different effects (Kalyuga, 2009). In learning environments that are not optimally designed, learners with high prior knowledge can compensate for this and still achieve good learning performance (Richter, Scheiter, & Eitel, 2018). In contrast, learners with low prior knowledge are more dependent on assistance and a design conducive to learning to achieve good results (Richter et al., 2018). However, to use a support approach, learners might need a certain amount of prior knowledge (Seufert, 2003). In contrast, for learners with high prior knowledge, there may be interference between the help and previously existing knowledge or strategies, and this may result in detrimental effects on performance (Kalyuga, 2009). The *expertise-reversal effect* describes these interferences of support with strategies of learners, and in turn, with higher prior knowledge. This implies that help that is offered should always stem from the consideration of the learner's prior knowledge (Vogt et al., 2020). When considering the potential influence of learner's characteristics (e.g., prior knowledge) on the learning process in VR, little research exists to date. Thus, potential effects of support approaches are deduced from prior publications with conventional learning materials.

Transferring these findings to the context of VRLEs, the effect of prior knowledge can be expected to become apparent for different aspects. Learners with low prior knowledge might lack the experience to differentiate between relevant and irrelevant information in the VRLE (Vogt, Babel, et al., 2021b). As a result, these learners might benefit from signaling (van Gog, 2014). Learners with higher prior knowledge, on the other hand, can be biased by signals highlighting information that is irrelevant to them, which they had already integrated into their previously existing mental model and should therefore rather focus on further unknown details (Richter & Scheiter, 2019). Learners with low prior knowledge have difficulty recognizing corresponding elements in the learning environment and linking them (Richter et al., 2018; Folker, Ritter, & Sichelschmidt, 2005). Here, elaboration prompts might activate learners to search for potential connections (Endres et al., 2017). Furthermore, adequate sequencing might help learners with low prior knowledge, as these learners lack a brief scaffold of the learning content in their mental model (Eitel & Scheiter, 2014; Ainsworth, 2006). Consequently, prior knowledge might impact the effect of different support approaches and should therefore be

2. Theoretical background and related work

considered as an important moderator, or at least as a covariate, within VR studies.

2.7.3. Further characteristics of learners

In general, other characteristics of learners than the ones outlined in the previous sections might also impact learning in VR (Lee & Wong, 2008). For example, one clear influencing characteristic constitutes spatial ability (Lee & Wong, 2014). Here, different effects are conceivable: In terms of the *ability-as-enhancer* hypothesis, learners with high spatial ability might particularly benefit from representation in VR as they have enough remaining cognitive capacity to deeply process learning content and form a coherent mental model (Huk, 2006; Mayer & Sims, 1994). Furthermore, it is conceivable that spatial ability acts in line with the *ability-as-compensator* hypothesis. Learners with very low spatial ability could benefit particularly from the situated three-dimensional representation in VR since these are directly shown the content in a corresponding spatial configuration and thus do not have to mentally construct it themselves (Huk, 2006; Mayer & Sims, 1994). However, the importance of spatial ability for learning success also depends on the chosen learning content. When spatial configuration play only a minor role in the chosen learning content, a large effect of spatial ability on learning success is not necessarily to be expected. In this case, other abilities should be investigated for their potential influence on learning outcomes.

Lee and Wong (2014), for example, describes that learners with low spatial ability showed higher learning outcomes in a non-immersive VRLE compared to a conventional learning setting, while these differences were not found for learners with high spatial abilities. In general, spatial ability is assumed to be a primary cognitive factor in learning, which is especially evident in computer-based learning environments (Huk, 2006). Depending on the level of spatial ability, learners benefit from interactive animations or simulations (Höffler & Leutner, 2011). This is attributed to the fact that spatial ability influences the extent to which relevant information is extracted from animations. Furthermore, spatial ability is assumed to influence the extent to which learners can reconstruct the learning content later on. In addition, the integration of the information into the existing knowledge network might depend on spatial ability (Lee & Wong, 2014).

When using VR animations in combination with narrations or short textual labels, verbal ability might also have an impact on learning outcomes (Wallen, Plass, & Brünken, 2005). Verbal ability is thought to be related to general crystallized ability and thus might have a significant impact on learning (Cattell & Horn, 1978; Hakstian & Cattell, 1978; Winn, 1982; Kollöffel, 2012). However, although closely related to knowledge, the construct of verbal ability is still distinguishable (Schipolowski et al., 2014). Furthermore, verbal ability is related to working memory capacity and thus learners with high verbal abilities might have more resources to process the learning material (Conway, Kane, & Engle, 2003). In line with the *ability-as-enhancer* hypothesis, learners with higher verbal ability might perform better and might also experience advantages when answering post-test questions (Wallen et al., 2005). Furthermore, when they are asked to rate concepts for connectedness, they might do so by deducing potential relations between words (Schipolowski et al., 2014). Therefore, when investigating learning in

VR, verbal ability should additionally be considered as a covariate.

So far, three of the four factors of Jenkins (1979) were described in detail. Hence, the fourth factor, the assessment of learning outcome, will be described in more detail in the next section.

2.8. Assessment of learning outcomes

Before investigating different methods to measure learning outcomes, it is first important to understand what the result of successful learning is. After learning successfully, learners should be able to recall important definitions, understand the consequences of manipulating one component onto another, and apply the learning content flexibly to other contexts (Bloom, 1956). Hence, learners build a coherent mental model by integrating the new information into their pre-existing knowledge network (Schnotz & Bannert, 2003). To be able to further differentiate whether deep learning occurred or whether learners did not have sufficient resources to process the content semantically, the concept *mental model* must first be further defined and characterized.

2.8.1. Nature of mental models

A variety of definitions and properties of mental models are described in the literature. For example, Craik (1943) describes the term *mental model* as a *small-scale model* that reflects an external reality. This model allows one to anticipate situations based on past events or to test various hypotheses. Such mental models can enable learners to be more competent in responding confidently and efficiently to situations that arise (Johnson-Laird, 1981; Craik, 1943). In addition to this basic functionality of mental models, a consensus about the analogical nature of mental models can be described (Johnson-Laird, 1996). While these definitions allow us to draw conclusions about the basis of future actions, they do not provide any information about the nature of the mental models themselves, nor do they describe the different stages of mental model development. Rouse and Morris (1986) elucidated this in more detail by classifying mental models as a special form of knowledge. Here, expert knowledge is contrasted with general knowledge or superficial understanding of the subject matter. In contrast, White and Frederiksen (1986) described mental models from a different perspective. Here, mental models are viewed as structures intrinsic to knowledge. These structures allow an additional differentiation of different types of knowledge: declarative, procedural, and a control structure (White & Frederiksen, 1986). Fittingly, the description of mental models by Kraiger, Ford, and Salas (1993) also referred to the structure of the mental model itself: Here mental models are described as the context and organization of the information contained and previously processed. Thus, the focus of this definition is the understanding of a mental model as a knowledge network. It is emphasized that new and old knowledge must be connected to form a coherent mental model (Kraiger et al., 1993). Based on these definitions, in the present thesis, the term *mental model* is understood and characterized as follows:

2. Theoretical background and related work

Learners build a knowledge network based on the processed information presented in the learning environment. In this process, old and new information is linked and ideally stored in a coherent and comprehensive overall model in the long-term memory. Based on this model, learners can define basic concepts, correctly infer relationships and causalities between concepts, and transfer what they have learned to new contexts and apply it to other problems.

From the previous sections, the underlying cognitive learning processes that need to be managed successfully to build a coherent mental model have become apparent. But *what happens when the learners do not fully succeed in deeply processing the learning content?* To gain further insights, the next section highlights different developmental stages of mental models.

2.8.2. Stages of development of mental models

Based on Mayer et al. (2002) different stages of mental model development can be differentiated: The *component* and the *causal mental model*. They postulated that only when learners face all challenges while learning successfully, can causalities between different concepts be deduced. Otherwise, the component and the causal mental model do not simultaneously build on each other due to the insufficient resources of the learner. This results in a more basic mental model: the component model. This basic mental model is reduced to a more local understanding of the concepts (Mayer et al., 2002). In this component model, the individual concepts are understood as a unit. This unit is labeled with a term. In addition, the contained concept is defined by certain properties or states. For example, based on such a component model, a learner can define the term *service robot* and also reflect facts about its degree of automation. However, complex comparisons with other types of robots or deductions for consequences, for example in the development of such a robot, are not possible based on the component model. To be able to describe the causal relationships in a topic and to derive more complex consequences, a causal mental model is required. Although an intuitive understanding suggests that the component model is created first and then extended, Mayer et al. (2002) assumes that the construction of both models, component and causal, occurs simultaneously. In the next section, measurement of different substages of processing the learning content are described as well as how different stages of mental model development might be deduced.

2.8.3. Measuring mental models

Based on the presented theoretical assumptions, two aspects of mental models can be distinguished: the functional aspect that is related to learning performance, and the structural aspect that reflects connections between concepts in the developed mental model. To better understand the functionality of the mental model, one can distinguish between different levels of learning outcomes that reflect whether the learning content was processed subsemantically or semantically. Furthermore, the structural assessments of the included knowledge network might deliver further information regarding how sophisticated the mental model is. Although

learners might not be able to build a correct mental model including causal relations between the entities, they might at least have a brief scaffold of concepts that are more or less related. These different perspectives on results from successful learning are related to different cognitive subprocesses.

Taxonomy of Bloom

To build a causal mental model, the relevant learning content needs to be selected, organized, and integrated. Hence, learners first process the learning content syntactically or superficially. This level of processing reflects learning outcomes on the knowledge level (Bloom, 1956). Learners can define basic concepts by repeating the given definitions and facts of the learning content. On this level, they do not need to understand the concepts, as simple recall of information is sufficient to answer knowledge questions. When processing the learning content semantically, learners can deduce potential connections between the different concepts and conclude consequences of manipulating different components. Furthermore, they can contrast two concepts, although this comparison was not explicitly included in the learning material. This implies that learners have developed a brief causal mental model that allows for the deduction of basic connections including directed relations between concepts. This is referred to as a learning outcome on the comprehension level (Bloom, 1956). However, on this level, learners are not able to understand more global connections between the concepts, and thus, although having processed the learning content semantically, they lack the flexibility to apply the new concepts to different tasks and settings. Some aspects of the comprehension levels might already be explained based on propositional networks which are still not analog and thus, the included information must be translated before being integrated into the mental model (Schnotz & Bannert, 2003).

To apply the learning content or to answer questions on the application level, learners need to process the content semantically and integrate the new information into the analog mental model. This enables learners to deduce more global relations between concepts, including correct cause and effect chains with more than two concepts. An analog and coherent mental model is thereby needed to apply and transfer the learning content to new contexts or problems.

Structural assessment of knowledge

As the taxonomy of Bloom (1956) is a functional performance measure of the mental model, it states little about the underlying knowledge structure. However, as pointed out by Kraiger et al. (1993), the underlying knowledge structure of mental models is also an important indicator of the extent to which a complex mental model has been formed. One method to uncover such immanent structures of the mental model is the so-called *Structural Assessment of Knowledge* (SAK). This is a psychometric scaling method based on graph theory and has already been applied in the context of learning. An important assumption is that declarative knowledge is relational (Kraiger et al., 1993). Thus, individual concepts should not be considered in isolation from one another. This structural approach aims to display information about mapping in

2. Theoretical background and related work

a complex knowledge network (Davis, Curtis, & Tschetter, 2003). Using this approach, it is possible to measure the extent to which specific concepts have been integrated into the knowledge network during learning, as well as how the corresponding mental model evolves. Thus, another facet of learning is mapped here: understanding similarities and relationships between concepts, definitions, or processes (Davis et al., 2003). Specifically, the SAK method involves the pairwise comparison of different concepts (Trumpower & Vanapalli, 2016). In each case, the connectedness of two concepts is rated. To be able to conclude about the development of the mental model, a pre-measurement is always compared with a post-measurement (for further details see Appendix A).

Using a multi-method approach

Using a multi-method approach by combining the functional and structural measurement approach offers the opportunity to analyze the developmental level, and thus, the quality of the resulting mental model in detail (Vogt, Babel, et al., 2021b; Makransky, Borre-Gude, & Mayer, 2019). Furthermore, this multi-method approach allows conclusions to be drawn about the relationships between the functional approaches to measuring learning outcomes and the structural perspective.

3. Research aims of the present thesis

The present thesis aims to gain further insights into the underlying cognitive learning processes while learning in VR. So far, no systematic findings are available regarding how different support approaches for learning in VR influence underlying cognitive learning processes. Additionally, this thesis reflects on possible interactions between learner's characteristics and the chosen support approaches. To clarify the different levels of learning outcomes and the different stages of mental model development, a multi-method approach is applied. This constitutes a unique feature of the present thesis. Although previous research in the learning context has considered functional and structural approaches respectively, such approaches have yet to be combined or contrasted in detail. Therefore, the present thesis does not only contribute to exploring the learning setting in VRLEs, but also to providing information into the development of mental models by use of a multi-method approach to investigate learning outcomes. Furthermore, this thesis contributes to the research field by developing a theoretical model (EMIL VR) which is based on prior theoretical and empirical findings and illustrates the underlying cognitive learning processes involved in immersive VR.

3.1. Extended Model of Immersive Learning in Virtual Reality

As a theoretical framework for this dissertation, the *Extended Model of Immersive Learning in Virtual Reality* (EMIL VR) is postulated. EMIL VR is built on previous theoretical and empirical findings and is therefore rather comprehensive and complex. Thus, each component will be described in detail and the model will be built up section by section. The existing research gap and the resulting research questions are provided in the context of EMIL VR.

3.1.1. Cognitive processes while learning in virtual reality

The underlying cognitive learning processes constitute the core of EMIL VR (see Figure 3.1). In line with the CTML (Mayer, 2005) and the IMTPC (Schnotz & Bannert, 2003), different cognitive subprocesses are included. Learning content is first selected for the presentation formats in the VRLE. In the next step, the information is organized in the working memory and processed semantically. In the end, new information is integrated into the mental model. The processing is assumed to depend on the chosen presentation format: The analog nature of pictures allows for the inclusion of information more directly into the mental model, while symbolic, textual information needs to be parsed before the integration (Schnotz & Bannert, 2003). Another important aspect of the included cognitive learning processes is the shortcut between the internal, text surface representation and the mental model (Schnotz & Bannert,

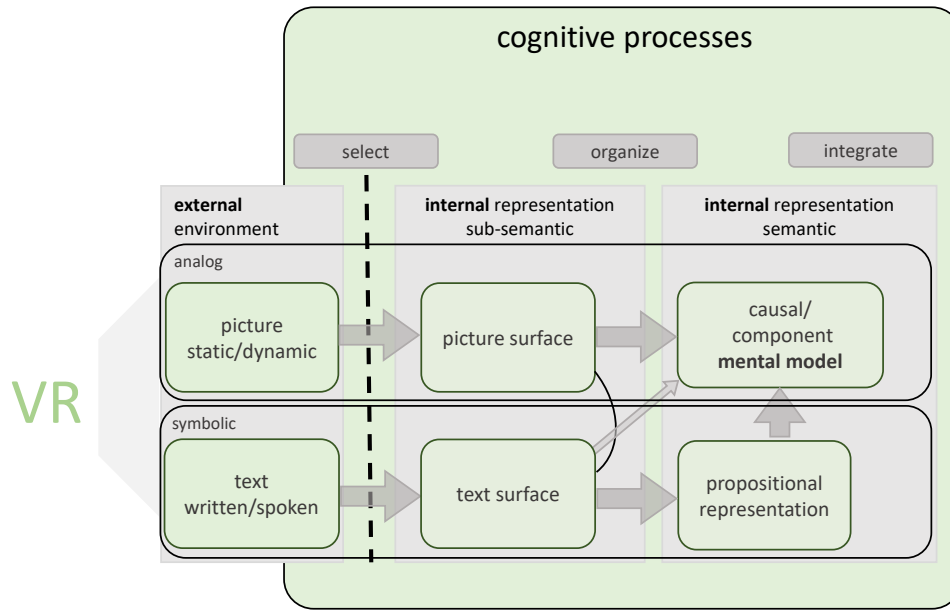


Figure 3.1.: Underlying cognitive learning processes while learning in virtual reality included in Extended Model of Immersive Learning in Virtual Reality

2003). As previously outlined, in VRLEs, learners might lack appropriate learning behavior. This might lead to superficial processing and thus the triggering of this shortcut (Vogt, Babel, et al., 2021b). Furthermore, it is assumed that different developmental stages of the mental model can be distinguished. In line with the distinction of Mayer et al. (2002), causal and component mental models are included as two potential stages of mental model development. Superficially processed learning content should rather lead to the formation of a component mental model. For deducing the correct cause and effect chain between the different concepts, semantic processing would be necessary.

3.1.2. External and internal support approaches for learning in virtual reality

As the next element, EMIL VR contains different support approaches that might substantially impact the underlying cognitive learning processes (see Figure 3.2).

The two support approaches are distinguished as being external and internal, respectively. As previously discussed, external support is related to the learning material that was postulated as being an important factor by Jenkins (1979; F2). Hence, this support approach refers to the specific design of the learning content and the chosen presentation formats. Depending on the chosen formats, as well as the combination of these formats, different underlying cognitive processes are triggered. Based on previous theories and empirical findings, two external support approaches were chosen to overcome potential challenges while learning in VR:

The first external support approach constitutes adding annotations. Annotations can be expected to support learners in finding relevant information in the VRLE and to help learners integrate information in the mental model. The second external support approach focuses on reducing strain to build a mental model. By adequate sequencing of the learning

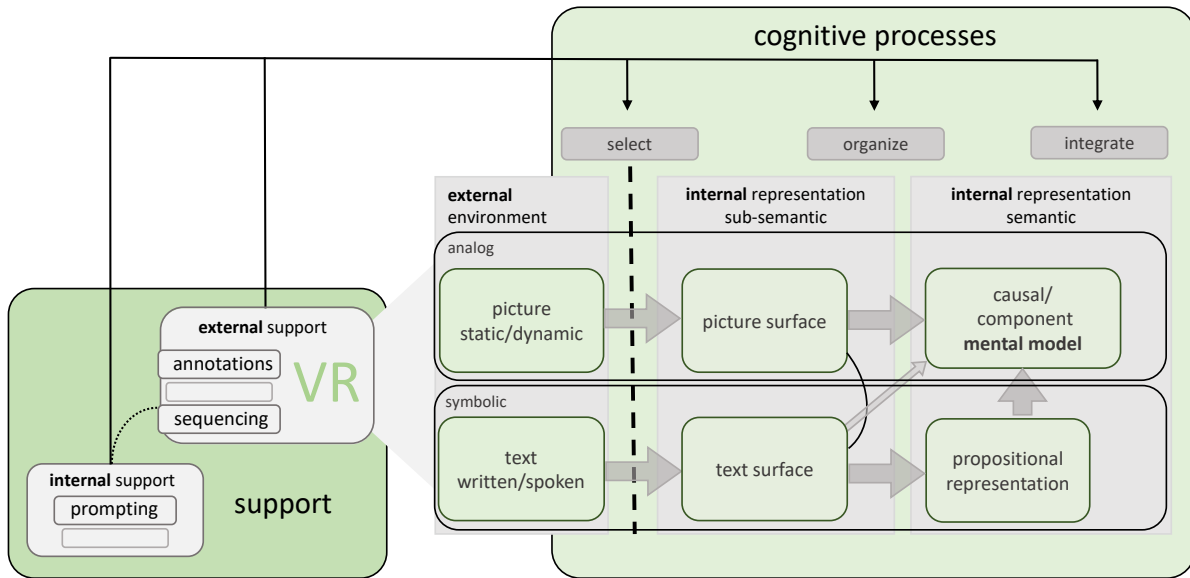


Figure 3.2.: Internal and external support approaches included in Extended Model of Immersive Learning in Virtual Reality

material, learners should have enough cognitive resources to process displayed learning content semantically. Consequently, they should be able to build a coherent mental model. By providing a brief pictorial overview, followed by a more specific text, learners may be additionally supported to build a causal mental model. These two support approaches in particular were selected as they seemed to be especially promising in overcoming potential barriers while learning in VR. Nevertheless, there are more external support approaches that also hold promise for supporting learning in VR. Such additional external support approaches are not yet included in EMIL VR. However, in line with this, the notion that EMIL VR should be further extended is indicated by the inclusion of an empty box in the external support box (see Figure 3.2).

As previously emphasized, learners can experience challenges in exhibiting adequate learning behavior in VRLEs. This is especially true as for most learners, such learning environments are rather unfamiliar. As highlighted by Jenkins (1979), in the factor learning setting (F1), specific instructions that are provided prior to learning can substantially impact cognitive learning processes. Thus, learners can be expected to engage in adequate learning behavior and to semantically process learning content. To this end, activating learners by utilizing prompts is one promising internal support approach. Depending on the chosen prompt, learners might be supported in different cognitive subprocesses. A major challenge in this regard, however, constitutes the deep processing of learning content. Hence, using elaboration prompts that trigger the deep processing of learning content is a promising approach to overcome superficial processing. In line with this, the internal support approach of prompting was thereby implemented into the present dissertation, while at the same time, it is theoretically reasonable that further internal support approaches can be added to EMIL VR in the future.

As the first leading question of this thesis, the impact of the chosen support approaches on learning outcomes was considered.

3. Research aims of the present thesis

Q1 Are there positive effects of the chosen support approaches for learning in VR?

As the present thesis focused on the different subprocesses while learning in VR, the first main research question was additionally extended to ask which level this potentially beneficial effect of the chosen support approach becomes apparent on.

Q2 On which subprocess do the positive effects of the chosen support approach become apparent?

As previously discussed, characteristics of learners play an important role in VR learning. Thus, as recommended by Jenkins (1979), this factor was additionally included in EMIL VR.

3.1.3. Critical learner's characteristics for learning in virtual reality

Based on previous models and empirical findings, two presumably important learner's characteristics are focused on in this thesis. Specifically, EMIL VR contains the two important learner characteristics of intrinsic motivation and prior knowledge (see Figure 3.3).

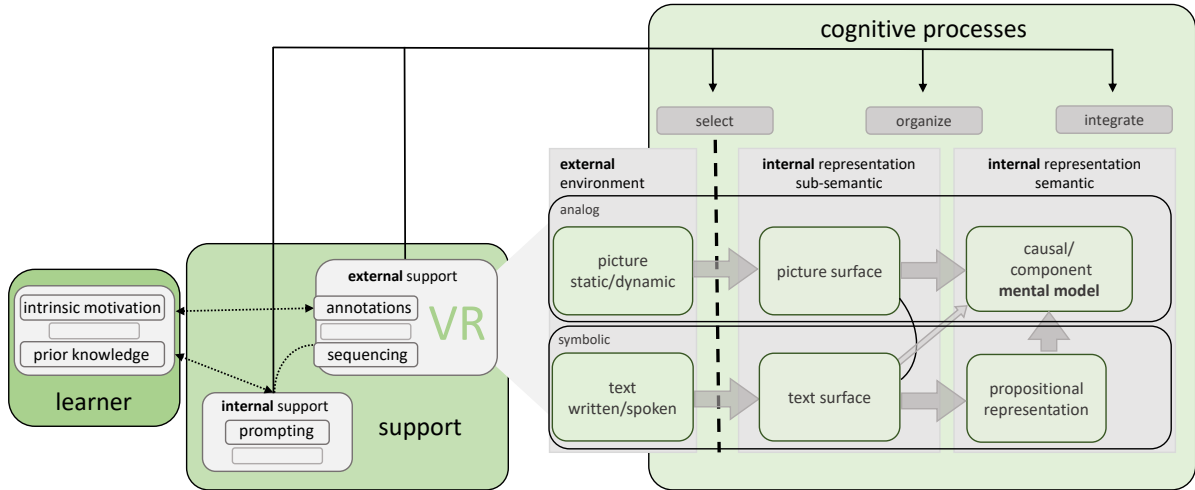


Figure 3.3.: Learner's characteristics included in Extended Model of Immersive Learning in Virtual Reality

As previously discussed, learners face the challenge of selecting relevant information in VRLEs. One characteristic that influences learner's subprocesses while learning is intrinsic motivation. Highly intrinsically motivated learners are better at selecting relevant information in the environment (Miranda & Palmer, 2013). Using certain support approaches, such as annotations, might interact with the intrinsic motivation of learners. Both adding annotations and high intrinsic motivation influence the same subprocess while learning: selecting the relevant information in the VLRE (Vogt, Albus, & Seufert, 2021). This interaction should consequently be considered when investigating the beneficial effect of adding annotation in the VRLE.

Furthermore, one important characteristic of learners is prior knowledge. In conventional learning environments as well as in VRLEs, the way that learners select information as well as the subsequent processing might be substantially influenced by their prior knowledge. Learners

with higher prior knowledge tend to guide their attention to relevant information in the learning environment (Vogt, Albus, & Seufert, 2021; Miranda & Palmer, 2013). Hence, the challenge of selecting important aspects of the VRLE might induce strain for learners with low prior knowledge (Ragan et al., 2015; Mulders et al., 2020; Richter et al., 2018). Further information processing, such as organization in the working memory and integration into the mental model is also reliant on prior knowledge (Schnotz & Bannert, 2003; Mayer, 2005). Furthermore, whether a support approach for learning in VR is successful can additionally be impacted by prior knowledge. An interplay between support approaches and prior knowledge is thereby theoretically reasonable. Aside from these two exemplary learner's characteristics, other promising characteristics can be included in future research. This is again indicated by including an empty box (see Figure 3.3). Based on this component of EMIL VR, the following research question was investigated in this thesis:

Q3 Do learner's characteristics influence the effect of the chosen support approaches on learning outcomes?

3.1.4. Assessing learning outcomes after learning in virtual reality

As a final element, EMIL VR describes different perspectives on learning outcomes (see Figure 3.4). As recommended by Jenkins (1979), different approaches to assess learning outcomes (F4) and mental model development should be considered to clarify the underlying cognitive learning processes.

In this thesis, two general approaches were included: one functional approach assessing the learning outcome as a performance test and one structural approach assessing mental model structure by the pairwise rating of connectedness between concepts. The functional, differentiated performance measurement is based on Bloom's taxonomy (Bloom, 1956): knowledge, comprehension, and application. The three levels of Bloom (1956) are included depending on the processing of the learning material. The knowledge level is related to surface representations, while comprehension and application are related to semantically processed information. Some aspects of questions on the comprehension and application level might be answered based on the network including propositional representations (Schnotz & Bannert, 2003; Díaz & Ioannou, 2019; Bloom, 1956). To answer other parts of the questions on these levels, the integration of the new information into the analog mental model is needed (Schnotz & Bannert, 2003; Díaz & Ioannou, 2019; Bloom, 1956). The structural approach captured structural accuracy of pairwise ratings of concepts for their connectedness compared to expert rating via the *Structural Assessment of Knowledge*. Depending on the processing level, learners can correctly classify concepts and answer questions on the performance test. The resulting structural accuracy is related to the semantically processed information as well as the surface representation. As both application and structural accuracy are related to a more global knowledge network, these two criteria are assumed to be positively related.

Based on this, the question about the potential relation of the two aforementioned approaches of assessing learning outcomes arises:

3. Research aims of the present thesis

Q4 How are the different approaches to measuring learning outcomes related to each other?

The postulated EMIL VR was systematically tested in three empirical studies. The effects of different support approaches depending on learner characteristics on different subprocesses while learning in VR were particularly focused on in these studies.

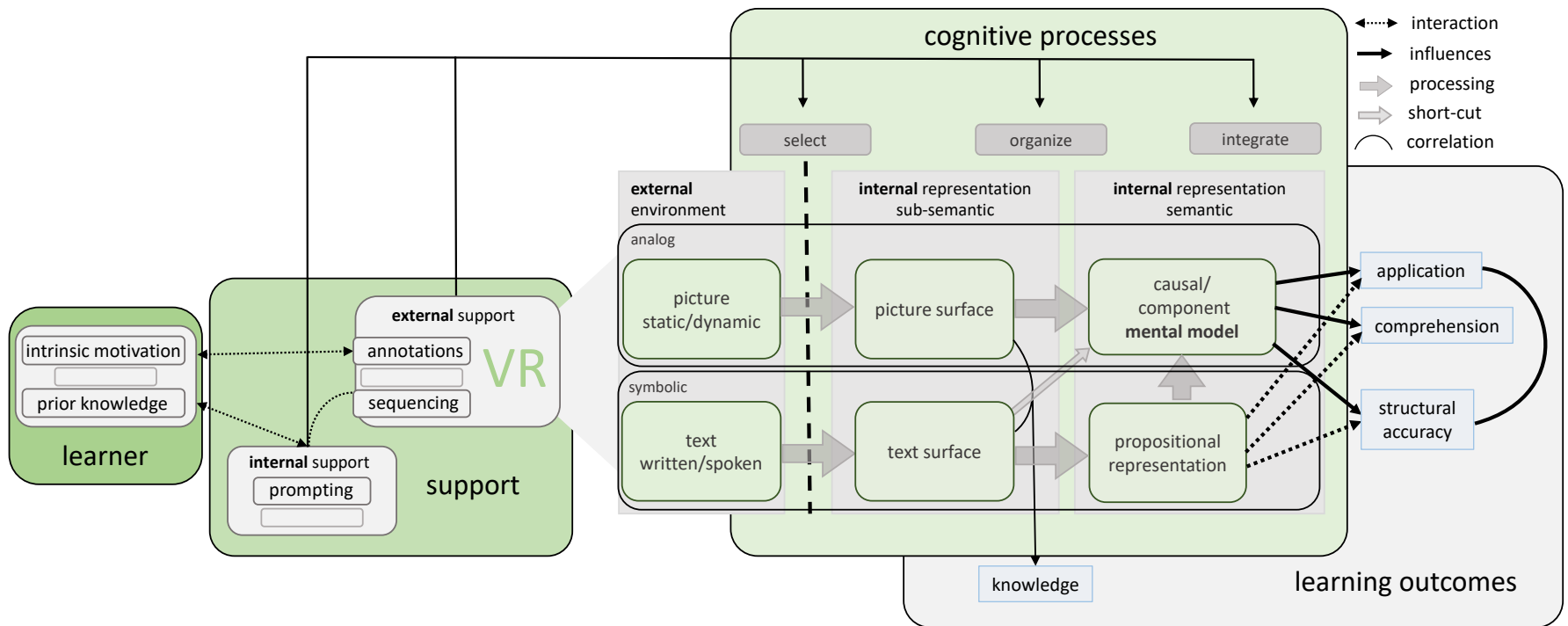


Figure 3.4.: Extended Model of Immersive Learning in Virtual Reality (EMIL VR)

Part II.

Empirical research

4. Overview of the three included studies

Within the theoretical background of this thesis, specific challenges while learning in VR were outlined. Based on this, different support approaches to overcome potential barriers to learning in VR were investigated in three empirical studies (see Figure 4.1). Stemming from the recommendations of Jenkins (1979), four important factors including the learning setting (F1), learning material (F2), learner's characteristics (F3), and assessment of learning outcomes (F4) were considered and included in the developed EMIL VR. Two external support approaches related to the specific design of the VRLE (F2), and one internal support approach related to the learner's active role and thus the learning setting (F1) were examined.

The first challenge discussed in this thesis was that VRLEs can be visually complex, and thus, the selection and integration of relevant information might be challenging (Makransky & Petersen, 2021). This challenge is related to the chosen learning material (F2) and the specific design of the VRLE. As a first external support approach, annotations (textual labels) were examined in Study 1 (see Figure 4.1). These annotations might support learners in selecting relevant information more efficiently and thereby help learners to map corresponding information.

Furthermore, given that for most learners VRLEs constitute a new learning setting, they may lack adequate learning behavior (Goodwin, Wiltshire, & Fiore, 2015). This challenge is related to the learning setting (F1) as well as the provided interaction activities to show respective learning behavior. As an internal support approach, and the second support approach chosen in this thesis, the active role of learners while learning was addressed. By using an elaboration prompt, learners were activated to engage in the learning process and to deeply process the learning content in Study 2 (Berthold et al., 2008; Vogt, Babel, et al., 2021b).

Moreover, visual complexity and presenting many details of the learning content simultaneously can lead to a high cognitive and perceptual load, which might result in less learning or superficial processing. As a consequence, learners may fail to develop a coherent mental model (Mayer et al., 2002). To overcome this barrier, Study 3 used a combination of external and internal support approaches (see Figure 4.1). As the second external support approach, adequate sequencing of the different presentation formats to present learning content was utilized in Study 3. By using sequenced material, this overload should be avoided. Adequate sequencing of pictorial and textual formats can be expected to support learners to build a coherent mental model (Eitel & Scheiter, 2014; Endres et al., 2017). As an internal support approach, an elaboration prompt was chosen to stimulate semantic processing. This in-depth elaboration also goes along with deducing connections between the different concepts and is thus assumed to be dependent on the pre-existing knowledge structure. Thus, combining both

4. Overview of the three included studies

approaches (internal and external) by adequate sequencing and prompting might result in a synergic effect.

As previously outlined, the underlying cognitive learning processes are dependent on learner's characteristics (F3). Thus, in the present thesis, important influencing factors were selected and examined based on previous theories and findings as well as their potential moderating effect on learning outcomes and underlying cognitive learning processes (Moreno & Mayer, 2007; Mayer, 2005; Schnotz & Bannert, 2003).

Potential moderation effects are examined in form of aptitude-treatment interactions. Both functional measurements related to learning performance and a structural approach to learning more about the configuration of the mental model are utilized and contrasted.

In the following a short summary, repeating the main theoretical aspects for each study, is included. The full version of the three peer reviewed journal articles can be found in Appendix A.

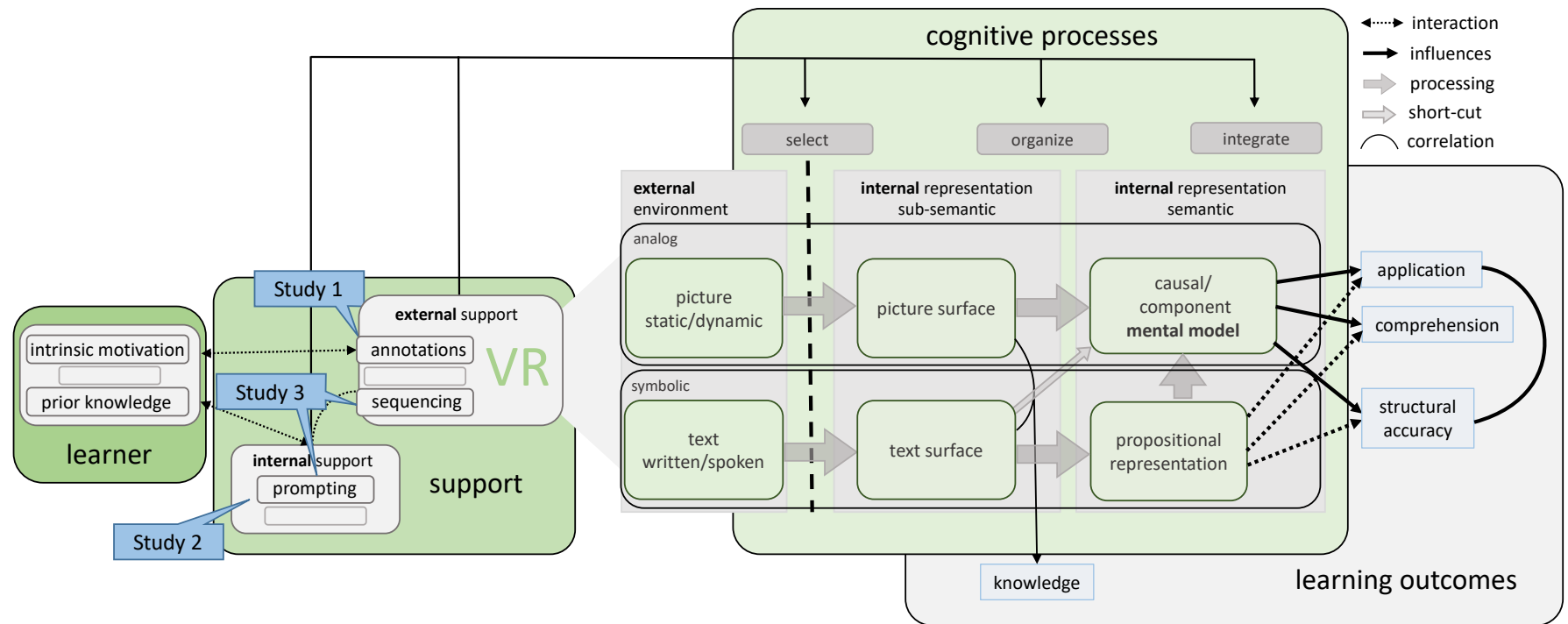


Figure 4.1.: Advanced organizer based on the Extended Model of Immersive Learning in Virtual Reality and the three experimental studies

4.1. Study 1 - Learning in Virtual Reality: Bridging the Motivation Gap by Adding Annotations

4.1.1. Theoretical background

Learning complex or abstract scientific content can pose challenges for learners to process, select, and integrate relevant information. This is particularly true as such content is often not intuitive to understand and cannot be easily connected to knowledge from everyday life (Rapp, 2005). VRLEs offer the possibility to display micro and macro phenomena that would be not observable in a real laboratory setting (Potkonjak et al., 2016). As VRLEs impose a rather high visual load on the learners, adding certain elements might support learners to select relevant information and to map corresponding elements (Richter, Scheiter, & Eitel, 2016; van der Meij & de Jong, 2006). One approach to support learners is to include textual annotations into the VRLE. These annotations should support learners in connection to their signaling nature: Annotations might foster just-in-time processing of important elements of the VRLE by guiding attention and reducing unnecessary visual load (Richter et al., 2018; McTigue, 2009). Furthermore, annotations labeling the pictorial representation in the VRLE might ease the element-to-element mapping of textual and pictorial information, as these are presented just-in-place and therefore follow the recommendation to consider spatial contiguity when designing learning environments (Mayer, 2014; D. Gentner, Rattermann, & Forbus, 1993). In addition, annotations might repeat important terms and thereby foster learning, particularly on the knowledge level. However, learner's aptitudes or states while learning might have a substantial impact on learning outcomes. Particularly, intrinsic motivation might lead to a more efficient visual search pattern in the VRLE and may thus increase learning outcomes. To elucidate cognitive learning processes, different levels of learning outcomes were considered (Bloom, 1956). As the importance of aptitude-treatment interactions has been found in previous studies, we investigated the interplay of adding annotations as a treatment factor and learner's intrinsic motivation as an aptitude factor in this study.

4.1.2. Research questions and hypotheses

In this study, we focused on two research questions:

RQ1 Do annotations have a positive impact on learning outcomes?

In detail, we expected a beneficial effect of annotations compared to no annotations in the VRLE on all three levels of learning outcomes (knowledge, comprehension, and application; H1a). Based on the repetitive function of annotations, we expected the largest positive effect of annotations on the knowledge level (H1b).

In our second research question, we considered an aptitude-treatment interaction:

RQ2 Does learner's intrinsic motivation moderate the beneficial effect of annotations on learning outcomes?

We expected that intrinsic motivation to moderate the relationship between the treatment factor annotations and learning outcomes in the VRLE (H2).

4.1.3. Method and results

We tested $N = 61$ participants (34.43% male) with average age of 23.73 ($SD_{age} = 6.00$) in our between-subject VR study. The procedure is displayed in Figure 4.2.

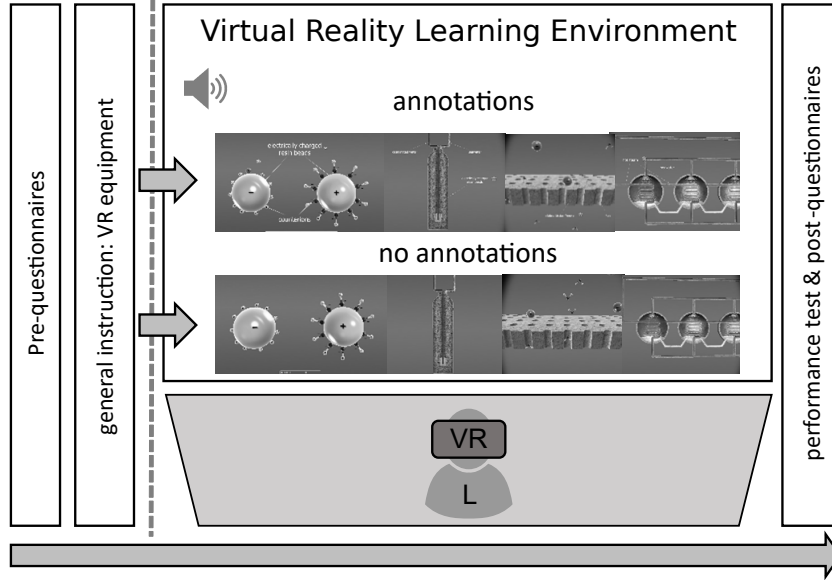


Figure 4.2.: Procedure of Study 1

Our participants were randomly assigned to one of the two different experimental conditions: they saw the VRLE with ($n = 33$) or without annotations ($n = 28$). As a dependent variable, we measured learning outcomes in a differentiated manner by distinguishing knowledge, comprehension, and application level (Bloom, 1956). In addition, learner's intrinsic motivation was assessed (Lepper et al., 2005; Harter, 1981). Based on theoretical considerations we measured and included prior knowledge as a control variable.

The expected beneficial effect of adding annotations into the VRLE on all levels of learning outcomes was not supported by the findings (H1a). However, based on the descriptive means of the two groups, the largest differences were found on the knowledge level, but as this result was not significant, the findings were not in line with our expectations (H1b). Considering the aptitude-treatment-interaction, we found that intrinsic motivation moderated the effect of adding annotations on learning outcomes. Further, follow-up analysis revealed that learners with intrinsic motivation of -0.7 or fewer points below the overall mean in particular showed beneficial effects of adding annotations on learning outcomes.

4.1.4. Discussion

In the first study, we uncovered the importance of considering learner's intrinsic motivation to deduce the effectiveness of design elements such as textual annotations to ease cognitive

4. Overview of the three included studies

processing of the learning content in a VRLE. Our results imply that learners are either intrinsically motivated or need to be supported by adding annotations to learn successfully in the VRLE. Different kinds of annotations or a different amount of annotations might further increase this beneficial effect. Furthermore, the displayed annotations might be extended and thus might support relation-to-relation mapping. This would provide further support to integrate the learning content. We focused on intrinsic motivation as an aptitude factor, however further learner's aptitudes might also explain a substantial amount of variance. Different types of VRLEs might be used to further investigate whether our results are transferable to VRLEs with more interaction possibilities.

4.2. Study 2 - Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model

4.2.1. Theoretical background

To present new technology, VRLEs offer great potential. However, using interactive VRLEs might lead to simply interacting with the environment without processing the presented learning content (Díaz & Ioannou, 2019). For instance, Mayer (2004) outlined that learners tend to explore a simulation instead of engaging in a learning process. Similar limitations of learning were reported when learning in a VR laboratory (Hu-Au & Okita, 2021). Thus, learners might need to be supported to show suitable learning behavior. One approach to stimulate learners to process the presented information semantically is to use an elaboration prompt (Endres et al., 2017). This elaboration prompt is an explicit hint for learners to process the meaning of the provided content and search for potential connections. Additionally, when including an explanatory component, this stimulates learners to translate the given learning content into their own words (Berthold et al., 2008). Hence, deep processing should be fostered by this prompt. Learner's characteristics might have a substantial impact on the effectiveness of prompting on learning outcomes (Roelle & Berthold, 2013). Thus, prior knowledge and verbal ability were both considered for their potential influence. To better understand the resulting mental model, two approaches were used to measure learning outcomes: one functional and one structural. The functional approach differentiates the levels of knowledge, comprehension, and applications (Bloom, 1956). The structural approach was based on the *Structural Assessment of Knowledge* to deduce the undirected connectedness of different concepts (Trumpower & Vanapalli, 2016).

4.2.2. Research questions and hypotheses

In this study, we focused on three research questions:

RQ1 Does prompting have a beneficial effect on functional learning outcomes?

In detail, a beneficial effect of prompting on all levels of functional learning outcome was expected (H1a). Due to the nature of the elaboration prompt, higher levels were assumed to show a larger beneficial effect (H1b).

In our second research question, the impact of prompting on structural learning outcome was issued:

RQ2 Does prompting have a beneficial effect on structural accuracy?

Prompting was expected to have a beneficial effect on post accuracy (H2).

In the last research question, the relationship of structural and functional measurement approaches was considered:

4. Overview of the three included studies

RQ3 Are structural accuracy and the application level of learning outcome related?

A positive relation between both measurements was expected (H3).

4.2.3. Method and results

We tested $N = 64$ participants (30 % male) with average age of 24.05 ($SD_{age} = 6.05$) in our between-subject VR study. The procedure is displayed in Figure 4.3.

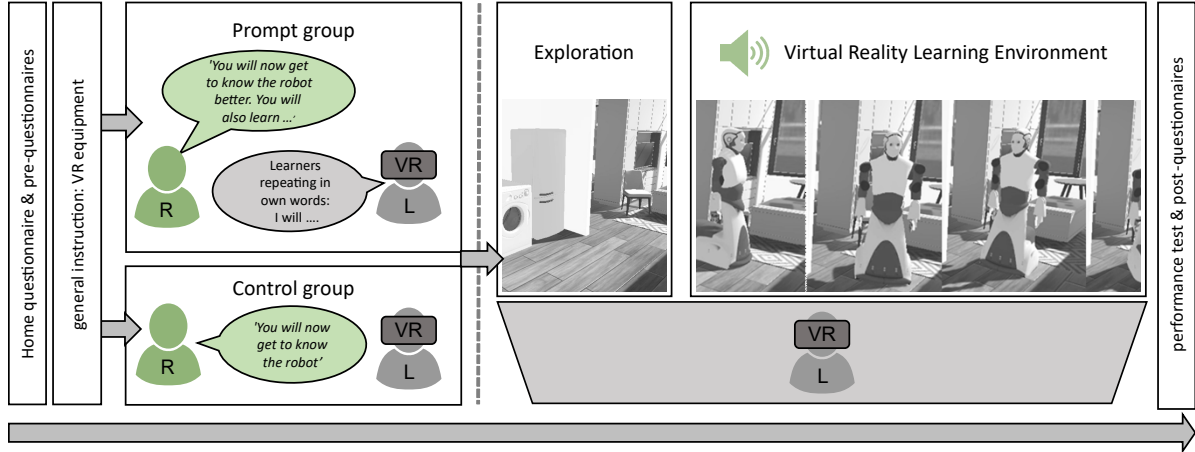


Figure 4.3.: Procedure of Study 2 (Vogt, Babel, et al., 2021b, p.7)

Our participants were randomly assigned to one of the two different experimental conditions: the learners in the experimental group received an elaboration prompt before entering the VRLE. This prompt pointed out that the VRLE contains information about a service robot and important concepts of robotics. Learners were particularly encouraged to connect the provided information and to prepare for a peer explanation after the learning unit. In contrast, the control groups were merely instructed that the VR unit was about to start and that they will see a robot. As a dependent variable, we measured functional learning outcomes by distinguishing knowledge, comprehension, and application level (Bloom, 1956). Additionally, structural accuracy was assessed before and after learning. Based on theoretical considerations, we also measured prior knowledge.

The expected beneficial effect of prompting before entering the VRLE was not found for all levels of learning outcomes (H1a). Solely a beneficial effect of prompting on the knowledge level was found. Consequently, the higher levels of learning outcomes did not benefit from prompting (H1b). Structural accuracy was increased by prompting when considering the pre-accuracy (H2). Application level and post structural accuracy were positively related (H3).

4.2.4. Discussion

In the second study, the importance of using a multi-method approach to investigate the internal support approach of prompting in virtual reality learning was discussed. The functional approach showed a positive effect of prompting on knowledge level and thus, superficial

4.2. Study 2 - Prompting in-depth learning in immersive virtual reality

processing was stimulated. Based on the structural accuracy of the mental model after learning, prompting stimulated the processing of learning content not only superficially. Moreover, a positive relationship was found between the functional application level and the structural accuracy of the mental model. This finding suggests that there may be further sublevels of mental model development and that shortcuts may be very important in processing learning content in VR. Future studies might combine further measurement approaches to gain insights into the mental model development.

4.3. Study 3 - Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome

4.3.1. Theoretical background

To overcome superficial processing while learning in VR, different support approaches are available. Learners experience the challenge of processing the learning content in VR deeply, as their cognitive capacity is limited and they can only perceive and process a certain amount of information per time (Mulders et al., 2020). Particularly, when learning with animations, which is not self-paced, this might challenge learners to process all provided information (Lowe, 2003). One approach to overcome this challenge is to reduce the information provided per time. This can be done by splitting up the learning material. For instance, the VR animation could be displayed before further details are provided as a narration. Thus, learners might have more capacity to process the provided learning content. However, this goes along with the question of which sequencing order should be used. Presenting a VR animation as a brief overview that is followed by a more specific narration is the preferable sequencing order based on previous findings (Eitel & Scheiter, 2014; Arndt et al., 2019). However, sequencing the material might challenge learners to integrate the information of different presentation formats (Mayer, 2014). Providing an elaboration prompt might support learners to connect and integrate the presented content (Endres et al., 2017). Hence, combining both support approaches, adequate sequencing and prompting, might result in beneficial effects and might overcome the challenge of insufficient cognitive resources.

4.3.2. Research questions and hypotheses

In this study, we focused on the following research questions:

RQ1 Does adequate sequencing have a beneficial effect on functional learning outcomes?

In detail, presenting VR animation first followed by the narration was expected to increase learning outcomes (H1). Additionally, the main effect of prompting using the sequenced learning material was examined:

RQ2 Does prompting have a beneficial effect on functional learning outcomes?

The elaboration prompt was assumed to increase learning outcomes on all levels (H2a). Additionally, the largest beneficial effect was expected on the highest level of learning outcomes (application; H2b).

As both support approaches were combined, the third research question was:

RQ3 Is there a synergic effect of both support approaches on learning outcome?

In detail, the highest learning outcome was expected in the group with VR animation and prompting compared to all other groups (H3).

4.3.3. Method and results

We tested $N = 81$ participants (37 % male) with average age of 26.08 ($SD_{age} = 6.71$) in our 2x2 between-subject VR study. The procedure is displayed in Figure 4.4.

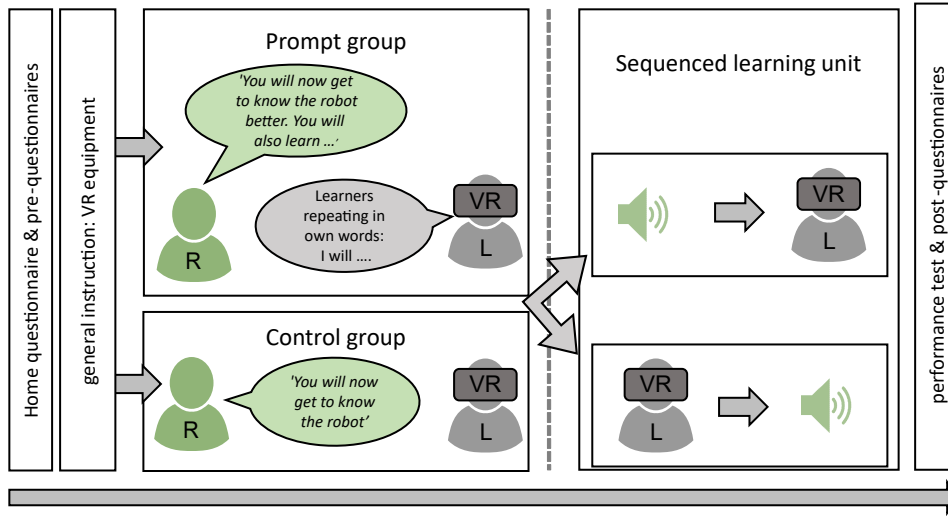


Figure 4.4.: Procedure of Study 3

The learners were randomly assigned to one of the four experimental groups of our 2x2 between-subject design study: with or without prompt and VR animation or narration first. In the prompt group, learners received the elaboration prompt before the learning content was presented. Depending on the sequencing condition, the VR animation or the narration was presented first. The learning outcome was measured by distinguishing knowledge, comprehension, and application level (Bloom, 1956). Critical characteristics of learners such as prior knowledge and verbal ability were assessed.

Sequencing had a beneficial effect on learning outcomes on the knowledge level (H1). The expected beneficial effect of prompting before entering the VRLE was not found for all levels of learning outcome (H2a). Nevertheless, a beneficial effect of prompting was found on the application level (H2b). No synergic effect of adequate sequencing and prompting was found (H3).

4.3.4. Discussion

In the third study, the importance of supporting learners while learning in VR was outlined. The two support approaches, adequate sequencing and prompting supported learning outcomes on different levels. While adequate sequencing helped learners to develop a brief understanding of the topic, prompting fostered semantic processing of the learning content. As the beneficial effect of the two support approaches unfolds on different levels, no synergic effect was found. Future studies might compare further combinations of integrated and sequenced learning material to uncover potential synergic effects with prompting. Furthermore, different external and internal support approaches might be combined to uncover the potential of these combinations.

Part III.

Discussion and implications

5. Discussion

After presenting the theoretical framework of this thesis by postulating EMIL VR and reporting the empirical results, in the following section, the main findings will be summarized and discussed. First, the results of the three studies are contrasted concerning the postulated leading questions (Study 1 – Vogt, Albus, & Seufert, 2021; Study 2 – Vogt, Babel, et al., 2021b; Study 3 – Vogt, Babel, et al., 2021a). Then, each support approach is discussed and EMIL VR is critically reflected on the empirical findings. Implications for theory, methodology, and practice are provided along with limitations and recommendations for future research on how to better support learners while learning in VR.

5.1. Main findings of the present thesis

In this thesis, the potential of VRLEs was outlined and specific challenges were derived regarding the characteristics of VRLEs. Based on these challenges, different support approaches (two external and one internal) were systematically investigated in terms of their potential for fostering deep learning in VR.

The first leading question focused on whether the chosen support approaches were beneficial for learning outcomes (Q1). To this end, both external and internal support approaches were found to have beneficial effects on learning outcomes. Hence, learners were successfully supported by adding the supportive elements (Vogt, Albus, & Seufert, 2021; Vogt, Babel, et al., 2021a, 2021b). However, not all approaches supported the learners to process the content deeply or semantically, and not all learners were supported in the same way (Vogt, Albus, & Seufert, 2021; Vogt, Babel, et al., 2021a).

As a differentiated measurement of the learning outcome was used in all three studies, and in Study 2, a multi-method approach to measure learning outcomes was used, the second leading question asked: Which level of learning outcomes was affected by the chosen support approach (Q2)? The results of Study 1 showed that only low intrinsically motivated learners benefited from adding annotations (external support) to process the learning content. Based on additional analysis (see Appendix B) the largest effect of adding annotations for learners with low intrinsic motivation was found on the comprehension level. In Study 2, the chosen elaboration prompt (internal support) supported learners to process the learning content superficially but did not stimulate semantic processing. Hence, the support approaches of both studies had limited effects as they did not foster learning on the application level.

In Study 3, two support approaches were combined: adequate sequencing (external) and the inclusion of an elaboration prompt (internal). Combining both support approaches was

expected to lead to synergic effects. However, the chosen support approaches were found to have beneficial effects on a different level of the learning outcome and thus, synergic effects were not observed. In contrast to Study 2, the results of Study 3 showed a beneficial effect of prompting on higher levels of the learning outcome. Based on the results from Study 2, which found that the beneficial effect of the prompt was limited on the knowledge level, the learning material was sequenced to conduct Study 3. In this study, the beneficial effect of prompting was found regarding semantic and deep processing of learning content (Vogt, Babel, et al., 2021a).

Reflecting on the results of Study 1, not all learners were supported by the addition of annotations in the same way. Potential moderating effects of learner's characteristics on the effectiveness of the chosen support approaches on learning outcomes were considered in the third leading question (Q3). Particularly, learners that scored substantially below the mean of the intrinsic motivation scale experienced a beneficial effect of adding annotations on the learning outcome. The importance of learner's characteristics was also reflected by the findings of Study 2. In this study, the beneficial effects of prompting on structural accuracy also depended on the initial levels of structural accuracy. Here, an expertise reversal effect was found for learners with high structural prior knowledge on structural accuracy after learning. Thus, both intrinsic motivation and prior knowledge had a significant impact on whether learners were supported, and additionally, to what extent learning was fostered.

By using a multi-method approach, the present findings shed light on the question of how different approaches used to assess learning outcomes are related (Q4). This question was investigated in Study 2. Here, a positive relationship between the learning outcome on the application level and structural accuracy after learning was found. This is a particularly interesting effect, as prompting only supported the learning outcome at the knowledge level, but no effects were found at the application level. Nevertheless, beneficial effects were found at the structural accuracy level, which were positively related to the application level, but not to the knowledge level. To better understand this complex pattern of findings, the results of the three studies are again examined using the postulated theoretical framework, EMIL VR.

5.1.1. External support approach for selecting and integrating learning content in virtual reality

The first support approach analyzed in this thesis involved adding annotations into the VRLE. Reflecting on the effect of annotations using EMIL VR, the findings from Study 1 were limited, as no overall main effect for annotations was discovered (Vogt, Albus, & Seufert, 2021). The findings suggest that only learners with lower intrinsic motivation experienced a beneficial effect from adding annotations. This superficial guidance of attention is particularly helpful for such learners, as intrinsic motivation is also a crucial factor for adequate attention guidance in the learning environment. Hence, learners had either an effective visual search pattern because they were intrinsically motivated, or they had additional help from annotations to optimize guidance of their attention (Miranda & Palmer, 2013; Vogt, Albus, & Seufert, 2021). This beneficial effect was found for the overall learning outcome (Vogt, Albus, & Seufert, 2021).

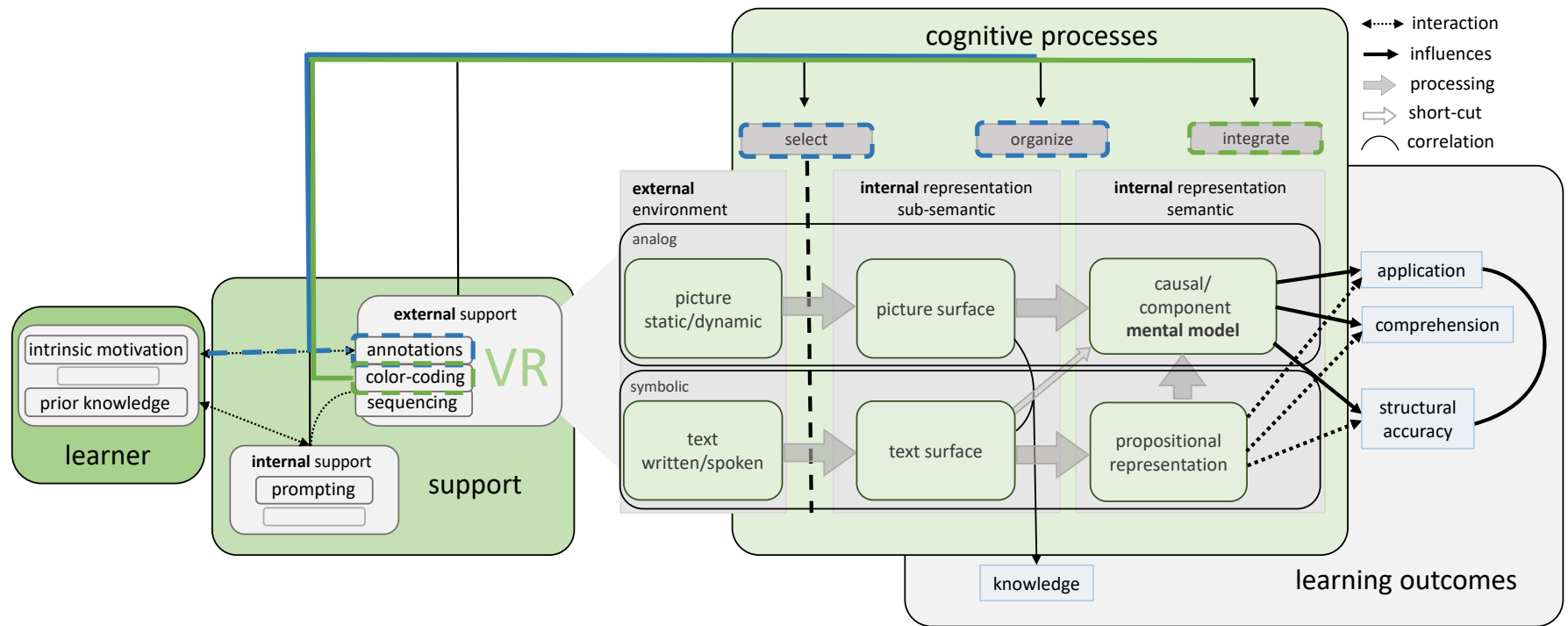


Figure 5.1.: Adapted version of the Extended Model of Immersive Learning in Virtual Reality

5. Discussion

When looking at the other results, this beneficial effect was limited, as bridging the motivation gap by adding annotations was not found for the application level (see additional analysis in Appendix B). Helping learners to guide their attention thereby does not necessarily guarantee deep processing or a flexible mental model. By directing attention, more processing time might be spent on relevant concepts. However, this does not guarantee that this information is processed semantically in a way that learners are able to apply the new information to another context and thus to answer questions on the application level (van Gog, 2014; Bloom, 1956). The annotations emphasized the central concepts of the learning content. They also repeated the content of the narration as a single word. Taken together, as annotations did not incorporate additional detail related to the learning content, information equivalency could be ensured (Noyes & Garland, 2008).

In the context of further research related to the present thesis, one study (additional Study 1; see Appendix B) examined the signaling effect by including color-coding of corresponding elements in a low immersive VRLE (Vogt, Albus, Montag, et al., 2021). In this study, color-coding eased the mapping of different presentation formats and directed attention to corresponding elements (see Figure 5.1). This study found a positive effect of color-coding on learning outcomes. In specific, as a result of color-coding, when controlling for prior knowledge, a beneficial effect was found on learning on the application level (Vogt, Albus, Montag, et al., 2021).

This is in line with previous findings (Richter et al., 2018; Vogt et al., 2020). Hence, although both support approaches (annotations and color-coding), were related to the same cognitive subprocesses, they fostered learning on different levels. Comparing the results of both approaches and incorporating a signaling nature, while the effect of annotations was limited on the lower levels of the learning outcome, the beneficial effects of color-coding were on the highest level of the learning outcome (see Appendix B). This finding was in line with previous results in that, when outlining corresponding elements, the mapping and integration of information is eased (D. Gentner, 1983; Fries, Son, Givvin, & Stigler, 2021). In contrast, annotations contain single words that repeat the content presented in the narration. This may encourage learners to process the learning content superficially and build a limited understanding of different relations between the concepts to answer questions on the comprehension level (Weinstein, Acee, & Jung, 2011). In this respect, although color-coding is less helpful as it points out connections between the presentation formats and directs attention but does not present any content (Renkl, 2005; Berthold & Renkl, 2009), it nonetheless showed beneficial effects on deeper processing, as the learning outcome on the application level was increased. However, this corroborated previous findings that outlined the beneficial effects of color-coding on the integration of different presentation formats (Folker et al., 2005). Furthermore, as annotations repeated the learning content, the support might shift to the knowledge level, as simply repeating content does not foster deep and semantic processing to build correct causal connections between the different concepts in the global knowledge network (Weinstein et al., 2011). While in the additional study with color-coding a main effect was found, in Study 1, only some learners were supported by the annotations.

Reflecting on the present results and the limited effect of annotations found in previous studies, several influencing factors can be considered in terms of their effectiveness in connection with adding annotations. Some studies show significantly larger effect sizes when stronger salience differences were present between the different groups (Faul, Erdfelder, Buchner, & Lang, 2009). Additionally, other variations of these annotations are conceivable (de Koning et al., 2010) and the number of labels used could be examined for answering the question of what kind of annotations are most helpful and which amount of annotations is best (McTigue, 2009; Vogt, Albus, & Seufert, 2021). Furthermore, different spatial placements could have replicated the design principle contiguity principle in VR (Ploetzner & Lowe, 2012; Mayer, 2014).

Additionally, the annotations could have been designed differently. In the learning environment, words were simply inserted without boxes or frames. On the one hand, this allowed the learners to still see the pictorial presentation formats, on the other hand, the annotations were not as salient as they would have been if they had also stood out from the environment in terms of color or illumination (de Koning et al., 2010). It was assumed that annotations also support the processing and therefore, the integration of the different presentation formats (Vogt, Albus, & Seufert, 2021; D. Gentner & Markman, 1997; Mayer, 2014). In addition, combining the external support approaches, i.e., adding annotations and color-coding, could support learners and improve learning outcomes at all levels. Further improvement of learning outcomes could be achieved by combining different external support approaches. Annotations and color-coding could be used to support the learners in certain parts of the learning material. Hence, future research might apply these results to systematically manipulate the annotations to deduce the relations between their characteristics and benefits on different levels of learning outcomes.

The lack of a main effect of annotations in Study 1 could also be due to limited test power. Descriptively, there was a tendency for higher learning outcomes in learning with annotations compared to the control group. However, the effect size in the present study is small, it therefore cannot be considered significant from an inferential statistics perspective (Faul et al., 2009). Nevertheless, this study outlined the importance of the active role of the learners: Learners who were either stimulated by external support or were intrinsically motivated to process the content deeply fared better in terms of learning outcomes. As a consequence of these findings, an important question constitutes whether this active role of learning can be promoted by using internal support approaches.

5.1.2. Prompting deep learning in virtual reality learning environments

In contrast to the external support approaches of adding annotations or color-coding, the elaboration prompt constituted an internal support approach related to learning behavior (Bannert & Mengelkamp, 2008). Based on previous findings, the chosen elaboration prompt aimed to trigger learners to process the learning content semantically (Berthold et al., 2008). In Study 2, a multi-method approach including both functional and structural approaches to measuring learning outcomes were combined. Reflecting on the results of Study 2 on the functional learning outcome, semantic processing was limited in both the prompt group and the control group. Hence, despite activating learners to process the learning content deeply, mostly

5. Discussion

superficial processing occurred. Nevertheless, the prompt showed beneficial effects on the knowledge level and a supportive effect was thereby found for processing the learning content sub-semantically. Looking at the effect of the prompt on the structural learning outcome, an interaction of elaboration prompt and pre-accuracy on structural accuracy after learning became apparent. This reflects that both the quality of their pre-existing knowledge network and the elaboration prompt influenced structural accuracy after learning.

Based on both functional and structural approaches to measuring learning outcomes, our findings allow deeper insights into the developmental process of the mental model while learning in a VRLE. The elaboration prompt enabled learners to increase their learning outcomes on the knowledge level and improved the structural accuracy. Therefore, we can assume that developing mental models occurs by building simple and undirected relationships as one additional developmental stage (Vogt, Babel, et al., 2021b). In the model of Mayer et al. (2002), two stages of mental model development were described: the component and the causal mental model. The component model entails a more local understanding of the concepts and their potential stages, while the causal model includes a complex knowledge structure with meaningful connections between the concepts. The present findings suggest a further substage, as learners build also connections between several of the included concepts (Vogt, Babel, et al., 2021b). However, they were not able to deduce correct causalities of these connections or to correctly connect all included concepts. As previously outlined, the development of a causal model requires sufficient resources and stimulation while learning (Mayer et al., 2002). In the present VR study, learners successfully built a brief component model that was extended by undirected connections but was unable to develop a more global and correct causal model (Vogt, Babel, et al., 2021b). By more closely examining the IMTPC, further details of the underlying cognitive learning processes in the VRLE can be deduced (Schnotz & Bannert, 2003). One important process for building a mental model with high structural accuracy is described as the shortcut in the IMTPC for including sub-semantic processed information into the mental model (Vogt, Babel, et al., 2021b). Our VRLE, in combination with the elaboration prompt, appeared to trigger processing via this shortcut. To develop meaningful relations between the different entities, additional semantic processing of the learning content is required (Schnotz & Bannert, 2003). Adding prompts enabled the learners to store the unspecific relations into the mental model but did not support the development of correct meaningful relations that would allow for answering comprehension and application (Vogt, Babel, et al., 2021b). Based on these findings, further research is necessary to better understand how to support deep learning in VRLEs.

Further activating approaches, such as prompts, have also been described in prior research to support learners with their learning behavior (Fiorella & Mayer, 2015). In the present thesis, the selected elaboration prompt was a rather minimal approach that could be further extended. For instance, prompting while learning has been described as being very efficient (Thillmann et al., 2009). In this thesis, the learning units were relatively short and it was assumed that the learners were already challenged by processing the learning content (Vogt, Babel, et al., 2021b). Hence, in the included studies, the prompt was provided before entering the VRLE.

However, including further prompts into the learning unit might support learners in processing the learning content more deeply (Thillmann et al., 2009).

As previously outlined, further types of prompts can be used to support the self-regulated learning process (Ifenthaler, 2012; Nückles, Hübner, & Renkl, 2009). In the present thesis, a highly pragmatic approach to prompt learners to process the learning content deeply was chosen. As this prompt can be applied flexibly and without the high cost or investing much time, it can be described as particularly economical, especially given that a significant positive effect on learning outcomes was already found using a small intervention.

Next to cognitive prompts, metacognitive prompts and motivational prompts are also important (Schumacher & Ifenthaler, 2021). Meta-cognitive prompts support learners by stimulating monitoring processes for learning (Berthold et al., 2007). For example, Pieger and Bannert (2018) used such metacognitive prompts in a hypermedia environment. They found a positive effect of these prompts on learning behavior. Furthermore, a positive effect of metacognitive prompts on learning outcomes was identified in previous studies (Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015; Daumiller & Dresel, 2019). Therefore, the use of meta-cognitive prompts seems promising in the context of VRLEs. Previous findings suggest that a combination of cognitive and metacognitive prompts can have a positive impact on learning outcomes (Berthold et al., 2007). This could be investigated in future studies in the context of VRLEs.

In addition, the importance of motivational processes was demonstrated by the findings presented in this thesis (Vogt, Albus, & Seufert, 2021). However, not all learners are intrinsically motivated to learn each topic. As a consequence, learners' motivation might be stimulated by adding supportive elements (Daumiller & Dresel, 2019). In literature, the use of motivational prompts is described as conducive to learning (Schumacher & Ifenthaler, 2021). Motivational prompts support learning by emphasizing goals or by stimulating the regulation of their motivation (Schumacher & Ifenthaler, 2021). Again, combinations with other prompt types or other support approaches are conceivable (Daumiller & Dresel, 2019).

5.1.3. Combining internal and external support for learning in virtual reality

In Study 3, both internal (prompting) and external (adequate sequencing) support approaches were combined (see Figure 5.2). As an internal support approach, the elaboration prompt developed for Study 2 was included in Study 3. While Study 2 presented the learning content simultaneously, in Study 3, the learning material was sequenced to overcome the limited effect of prompting on learning outcomes. This limited effect of prompting might have occurred due to insufficient cognitive resources to process the content deeply and to build a causal mental model (Mayer et al., 2002). To overcome this assumed cognitive overload, sequencing the learning material can be used as a strategy to reduce the information that needs to be processed simultaneously (Van Merriënboer & Sweller, 2005). Hence, the learners should have more resources to process each of the given presentation formats and should be able to prepare for later deep processing of the content (Johnson, Reisslein, & Reisslein, 2014).

As this finding is very well researched, it was not replicated in the present thesis. By

sequencing the learning material, the question of which presentation sequence would be more helpful for learners arose. Based on many findings and theoretical assumptions from multimedia learning with text and pictures, presenting the pictorial VR animation first, followed by the more specific textual presentation format, was assumed to be the adequate order to present the sequenced learning material (Arndt et al., 2019). When learning with adequate sequenced learning material, learners can build a brief scaffold based on the analog, pictorial information and add further symbolic details in the process (Eitel & Scheiter, 2014; Vogt, Babel, et al., 2021a; Arndt et al., 2019). Despite this strong evidence of presenting a brief pictorial overview first, based on other publications, presenting the pictorial presentation format after the more specific textual learning content might serve as an opportunity to repeat the learning content (Mayer, Bove, Bryman, Mars, & Tapangco, 1996). Additionally, by displaying the VR animation after the more specific text, learners can perceive a specific example and have the opportunity to apply the concepts of the learning content (Brunyé, Taylor, & Rapp, 2008). However, as prior theories and findings support that the adequate presentation order when sequencing is to present the pictorial presentation format first, this was referred to as adequate sequencing.

Despite the question of adequate sequencing order, sequencing learning material involves a further challenge. Although it was assumed that learners might have more cognitive resources to process the learning content, this might decrease the integration of the learning content from different presentation formats, when these are not presented simultaneously (temporal contiguity effect; Mayer & Fiorella, 2014). Previous findings outlined the benefits of presenting narration and pictures or animations simultaneously (Mayer & Anderson, 1991). In contrast, Arndt et al. (2019) could not replicate these findings and found no differences between the sequenced and the simultaneous presentation for the integration of the different presentation formats. They explained their finding by the fact that simple material was used that did not challenge the working memory capacity of the learners (Arndt et al., 2019). Consequently, when using more complex learning material, learners might need additional support to connect the different presentation formats and the learning content presented in the sequenced learning unit (Richter et al., 2016). In line with this idea, the elaboration prompt in combination with the adequate sequencing was assumed to have synergic effects on learning outcomes (Vogt, Babel, et al., 2021a). Building a mental model includes correct causal connections between the different concepts in the knowledge network. By processing further details and by mental model inspection, the resulting mental model can be further developed (Schnotz & Bannert, 2003). Hence, it is plausible that helping learners to build a brief scaffold based on an analog presentation format first that is refined by additional textual detail, can support learners. Adding an elaboration prompt might additionally activate the process of building a mental model. In contradiction to this expectation, the results of Study 3 did not reveal this synergic effect. To discuss this finding, the different effects of prompting and sequencing are added in EMIL VR (see Figure 5.2).

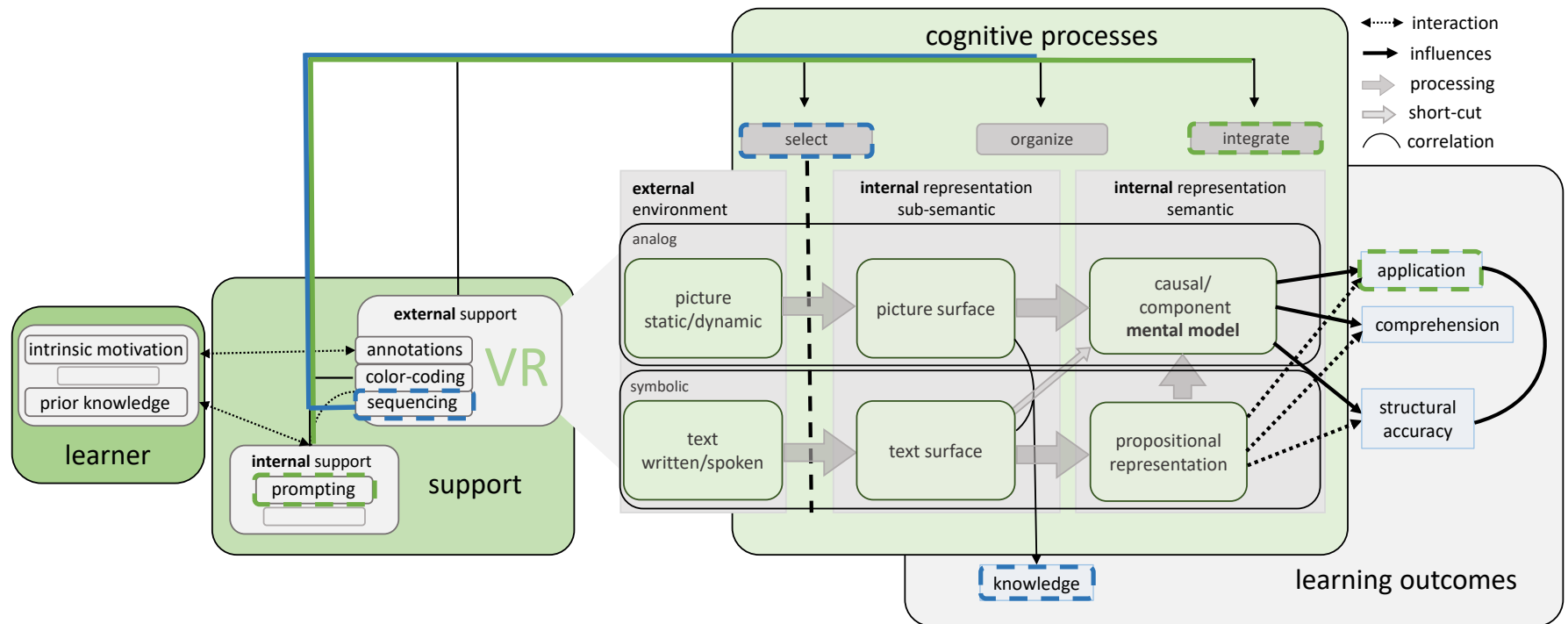


Figure 5.2.: Adapted version of the Extended Model of Immersive Learning in Virtual Reality

5. Discussion

In line with the marked subprocesses in Figure 5.2, the beneficial effect of adequate sequencing was found to foster superficial learning, while prompting supported learners to deeply process the learning content. Thus, as different subprocesses were stimulated by the external and internal support approaches, no synergic effect was found. The beneficial effect of the elaboration prompt on the learning outcome found in Study 3 is particularly interesting because in this study, unlike Study 2, the stimulation of deeper processing of the learning material was successful. Although the learning content in both studies was similar, the technical setup of the two studies was different due to pragmatic reasons. As this might cause bias, the integrated and the sequenced conditions cannot be used to conduct inferential analyses. Nevertheless, based on the present findings in the context of previous research, it can be assumed that the reduction of cognitive or perceptual load through adequate sequencing of the learning material made it possible for learners to process the learning content more deeply. As a consequence, the learners in Study 3 were able to answer questions on the application level more successfully due to the additional stimulation from the prompt. In the future, this assumption should be investigated in more detail. For instance, using a differentiated measurement of cognitive load including extraneous, germane, and intrinsic cognitive load might allow for additional insights into how the support approaches affect the available cognitive resources while learning (Klepsch et al., 2017; Sweller, 2011).

So far, we have focused on internal and external support approaches and their effects on learning outcomes. As outlined by Jenkins (1979) and supported by the findings of this thesis, the learner's characteristics should be considered in more detail when discussing the results. This is focused on in the next section.

5.1.4. Learner's characteristics and learning content

All three studies dealt with a complex interplay of many factors: the chosen support approach (internal or external), learner's characteristics, and the chosen assessment of learning outcome to uncover effects on the underlying cognitive learning processes.. As simultaneously considering these factors can lead to many potential interactions, each study focused on a select few as well as their consequences on learning outcomes.

In the present thesis, the findings support the assumption, that learner's characteristics might moderate or impact different support approaches (Vogt, Albus, & Seufert, 2021; Vogt, Babel, et al., 2021b). Two important characteristics were focused on, namely intrinsic motivation and prior knowledge. As expected from previous findings and theories, prior knowledge is crucial for learning outcomes and should be considered as a moderator or at least be included as a covariate to control for related variance (Tobias, 1994). The importance of intrinsic motivation on early processing of learning content was highlighted from the present results, and a beneficial effect of support was additionally found for learners with low motivation (Vogt, Albus, & Seufert, 2021). Moreover, the importance of prior knowledge, based on the assessment of structural accuracy, was found to be important for the supportive effect of prompting (Vogt, Babel, et al., 2021b). Prior knowledge was included for investigating all related research questions as a covariate. According to the present findings, the importance of including prior knowledge at

least as a covariate is emphasized (Vogt, Albus, & Seufert, 2021; Vogt, Babel, et al., 2021a, 2021b).

Furthermore, as the learning content about robotics conveyed learning content via textual presentation formats, the verbal ability was included as a covariate. In line with the basic assumption that learning is related to general cognitive abilities, verbal ability was found to be related to learning in VR (Pieger & Bannert, 2018). This variable should especially be considered when using learning material with written or spoken text (Schroeder, 2011).

As VRLEs offer the opportunity to display learning content including the spatial configuration of the included object, the learner's characteristics are likely to play an important role for learning in VR. Previous research on learning with animations uncovered heterogeneous effects of the impact of spatial ability on learning outcomes. While Mayer and Sims (1994), Huk (2006), and Lee and Wong (2014) assume the importance of spatial ability for learning with animations, Höffler and Leutner (2011) state that spatial ability does not necessarily play a major role in learning with animations. Therefore, spatial ability was investigated in the present research using a study that relied on the same learning material as Study 1. Here, principles of seawater desalination were presented including micro-processes such as chemical reactions. We found no significant effect of spatial ability on the learning outcome when using this learning material (Albus et al., 2021). Based on this result, it can be concluded that spatial abilities do not necessarily have an impact on learning outcomes in VR. Instead, only if the spatial component plays a central role in the chosen learning content will the spatial ability of the learners be an important factor influencing learning outcomes. For such content, the learners' spatial ability would be an important influencing factor for learning outcomes. The chosen learning material about seawater desalination (Study 1) as well as the learning material about robotics (Study 2 & 3) did not include spatial configurations that were substantial for understanding the presented concepts in the VRLE.

In addition to the chosen learning content, other factors have been described in the literature that can influence the effect of spatial skills on learning outcomes. For instance, in CAMIL a beneficial effect of agency on learning outcomes has been described, Merchant et al. (2014), however, did not uncover this beneficial effect for all learners. In their study, only learners with high spatial ability were found to benefit from high control and experienced beneficial effects on learning outcomes. Thus, spatial ability appears to be an important factor and should be examined in future studies. Here, the importance of the recommended factors by Jenkins (1979) becomes apparent. In this case, learning content and the level of control in the VRLE might be systematically manipulated to gain further insights into the underlying processes that cause this interaction and the effects on learning outcomes. So far, three of the four factors of Jenkins (1979) have been reflected and discussed based on the main findings of this thesis. The last factor, the assessment of learning outcomes (F4) will be addressed in the next section.

5.1.5. Assessing learning outcomes

To assess learning outcomes, two approaches were chosen: one functional and one structural. Although a multi-method approach was used to gain further insights into the developed mental

model, further approaches might be considered and might advance the field of research on learning in VR as well as the effectiveness of different support approaches.

Based on the results of Study 2, different developmental states should not necessarily be assumed as a dichotomy between the component model and the causal model as described by Mayer et al. (2002). Rather, there seem to be further gradations of the development process. Consequently, future research using multi-method approaches to measure mental models and learning outcomes might lend more information into this developmental process.

Furthermore, the functional approach to measure learning outcomes could be extended. The taxonomy of Bloom (1956) was used to differentiate between the knowledge, comprehension, and application level. The taxonomy was revised and further developed by Anderson and Krathwohl (2001). They extended the initial taxonomy that focused on cognitive processing of the learning content with additional dimensions by including metacognitive, procedural, conceptual, and factual learning outcomes. As the aim of the present thesis was to elucidate the cognitive learning processes, factual knowledge was focused on and levels of processing were differentiated based on the assumptions of the original taxonomy of Bloom (1956). Nevertheless, VR offers great potential to train and learn procedures or complex tasks (Makransky, Borre-Gude, & Mayer, 2019). VRLEs could be used, for example, to achieve learning goals beyond knowledge transfer, such as procedural knowledge or the training of very specific procedures and activities (Goodyear & Retalis, 2010; Makransky, Borre-Gude, & Mayer, 2019). This form of learning could also be well represented in VRLEs and drive knowledge transfer in a broader sense in many application areas (Makransky, Borre-Gude, & Mayer, 2019). For instance, when learning about chemical reactions, learners might learn basic concepts and more abstract principles combined with procedural learning content such as conducting a certain procedure in the lab. Both learning goals might be accomplished by using VRLEs. One additional advantage is that in VR, the spatial configuration of the learning content might ease the understanding of the relatedness of the concepts. In conventional settings, learners would be challenged to imagine these configurations and deduce the relations (Huk, 2006). However, measuring mental models and learning outcomes occurred by using the textual format. The structural approach intended to broaden insights into the development of the mental model, as the resulting mental model was described by the relatedness of two concepts. This approach involves a certain spatial component but still does not match the complexity of a 3D knowledge network. Future research might thereby focus on developing different measurement approaches to examine the mental model in a different way that is more similar to the used VRLE and allows for the measurement of the mental model in a more sophisticated way. Research indicates that different approaches to investigate the nature of the relations between the concepts exist (Trumpower & Vanapalli, 2016, e.g.). However, these are often very effortful as they are based on a qualitative analysis of the descriptions about relationships in the learning topic by the learner (Eckert, 2000). Again, these approaches are based on verbal information and learners need to translate their mental model to describe and not depict it. First approaches exist to use depictive representations to uncover the nature of the mental model as a drawn scaffold (Butcher, 2006). These might be extended to VR and enable learners to depict their mental model more intuitively.

5.1.6. Implications of the present thesis

In this thesis, EMIL VR was postulated as a theoretical framework model. This model is the first model to combine the four factors according to Jenkins (1979; learning setting, learning material, learner's characteristics, and assessment of learning outcomes) to systematically describe underlying cognitive learning processes in VRLEs.

Reflecting on the findings of the present thesis, promising internal and external support approaches have been included in EMIL VR. Firstly, the importance of considering the interplay of learner's characteristics and the support approach for investigating the learning outcome in VR was outlined. Next, additional developmental stages of mental models were deduced based on the included findings. EMIL VR was also extended in the course of discussion and further important interactions and influencing factors were highlighted. These points can be seen as a theoretical basis for future studies and can guide research focused on systematically examining the potential of VRLEs in the future. Adding to this, the aforementioned interactions of different factors should also be focused on in future research.

The necessity of measuring learning outcomes in a differentiated manner was supported by the findings of all three studies. Particularly, when addressing the question whether support approaches stimulate semantic processing, a differentiated measurement approach should be implemented. To measure learning outcomes and mental models in the present thesis, different approaches were considered to gain further insight into the cognitive learning processes that might be fostered by the chosen support approach. Both structural and functional approaches to measure learning outcomes and mental model development are valid methods to clarify how learning processes can be fostered. The present results imply that the functional and the structural approach show a positive relationship. However, the fact that the structural approach is only positively related to the application level provides more details into the development of the model. Furthermore, the particularly relevant cognitive sub-processes can be examined more closely based on this finding, especially when investigating the beneficial effect of prompting on different learning outcomes. The benefit of using a multi-method approach to better understand the underlying cognitive learning processes as well as the different stages of mental model development was shown. This central insight can be used in future research to systematically extend and empirically test the theoretical framework model EMIL VR.

In terms of implications for practice, the present research implies that learners generally benefited from the selected support approaches. Here it should be emphasized that when selecting a support approach, the learning content presented in the VR and the presentation formats should also be considered. Based on the knowledge gained from the present thesis about the supported cognitive sub-processes, specific support approaches can be selected. In summary, pragmatic and economical support approaches such as the elaboration prompt or the selection of the correct sequence of the presentation can already have beneficial effects on learning outcomes. The used support approaches had different effects on different levels of learning outcomes. Although this may be viewed as a limitation of the support approach, it can also be seen as a possibility to support learners according to their chosen learning

goal. Of course, the main goal of learning is always to develop a deeper understanding and a causal mental model. However, this cannot be achieved in every learning unit by every learner. Learners without prior knowledge cannot necessarily achieve this goal in a short time. Therefore, pragmatic support could first be provided on the surface, for example, by adequate sequencing or signaling by annotations. Then, in the following learning units, the knowledge network could be further linked by additional prompts, and causal connections could be built into this network. In general, previous findings have found that wrongly chosen or too much support can also lead to negative effects. Therefore, this aspect should also be considered in the context of VRLE (Kalyuga, 2009; Makransky & Petersen, 2021). Particular attention should be paid to the chosen support approach to reduce additional cognitive load, which can hinder successful deep processing. The learner's characteristics are of particular importance here, since the match of the support approach depends not only on the learning content, but also, as already mentioned, on prior knowledge and intrinsic motivation.

The expertise reversal effect described in the literature was found for prompting on structural accuracy. Learners with high structural prior knowledge seemed to suffer interference when the elaboration prompt was added. However, this pattern was not found on the functional level of the learning outcome or for the other support approaches. This could also be due to the pragmatic support approaches since they were chosen sparingly and from the beginning in such a form that they do not or hardly consume cognitive resources. Nevertheless, different variants of VRLEs could be individually provided according to learner's characteristics for optimal support. For example, after a short pre-survey, different VRLE variants could be selected according to expertise. It would also be conceivable in highly interactive VRLEs that, based on process data, learner's behavior could be used as an indicator of their relevant characteristics, thus creating an individual learning path. However, this requires further development of VRLEs in cooperation with different disciplines. In practice, it is important that teachers can also flexibly create learning content in VR and adapt it to their target group. One approach that seems promising here is in-situ authoring (Prouzeau, Wang, Ens, Willett, & Dwyer, 2020). Here, teachers can create the learning environment directly in VR. These authoring tools could then contain additional support for the teachers based on empirical findings, recommendations for support, and proven principles for the presentation of the learning content as examples or suggestions. In this way, the gap between research and practical application of learning in VR could be narrowed.

5.1.7. Limitations and further recommendations for future studies

The present thesis delivered many insights into the potential of learning in VR and how support approaches impact underlying cognitive learning processes and learning outcomes. However, many important questions concerning learning in VR remain unanswered. In this thesis, interaction effects of the aforementioned influencing factors were investigated based on previous findings. Nevertheless, further combinations of these factors are possible and need to be investigated in the future. For instance, instead of sequencing the presentation formats of VR animation and narration as separate parts, smaller units from the learning content could

be extracted to examine how a more complex sequencing of the learning content may impact learning outcomes. Furthermore, the spatio-temporal configuration of the VR animation could be manipulated, as the specific design of the presentation format might be pivotal for learning outcomes (Ploetzner & Lowe, 2012). Additionally, further complex interactions are theoretically reasonable. For instance, depending on the expertise level of learners, different sequencing of material could be used. Adapting the VRLEs as well as the chosen support approaches to the learner's characteristics might substantially impact learning in VR. Furthermore, based on Study 2, it was assumed that sequencing the learning material in an adequate order would lower the cognitive load and enable learners to process the learning content deeply. Based on previous research with conventional learning materials, many findings already exist that support and replicate this effect (Van Merriënboer & Sweller, 2005; Pollock, Chandler, & Sweller, 2002; Britton, Glynn, Meyer, & Penland, 1982). In the present thesis, this effect was therefore not investigated in more detail. However, for future research in the context of VRLEs, this finding of reducing cognitive load by sequencing the learning material could be systematically investigated with extended experiments.

Characteristics of the chosen virtual reality learning environment

As previously outlined, the chosen learning content presented in the VRLE has a substantial impact on whether learners might benefit from the in-situ learning setting and how they can be supported. In the present thesis, two exemplary learning units were utilized to investigate learning in VR. However, using learning material to present learning content with a stronger spatio-temporal dimension might offer additional insights into the potential of VRLEs and how to foster better learning outcomes. The chosen topic might interact with the effectiveness of the support approaches and the learner's characteristics. Furthermore, when choosing VRLEs to provide learning content, emotional factors might also be considered where VR could be used display certain situations to induce strong emotions. For instance, one study used VR to provide details related to war history. The resulting emotions when presenting this content in a situated and more realistic way should be considered, and topics should be chosen cautiously in order to minimize potential emotional strain (Calvert & Abadia, 2020).

One limitation of the present thesis was that rather short learning units were used. This is in line with the tradition of multimedia research, for instance, described and conducted by Mayer (2020) who published many empirical studies using learning units that were around two minutes long. Using longer learning units might challenge learners in different ways. By critically reflecting on their learning experiences and their development, they might optimize their learning behavior through planning, monitoring, and regulating. The importance of such monitoring processes might increase when using longer or repeated learning units in VRLEs (Howard & Lee, 2019).

The available degrees of freedom in the VRLE might impact learning outcomes (Makransky & Petersen, 2021). Both immersion and agency are related to the chosen setup of the VRLE and the possibilities to interact with the learning content (Makransky & Petersen, 2021). The interactivity of the VRLEs could be further extended and the effects of agency and immersion

on learning outcome might be further investigated.

Another promising possibility to learn more about how to support learners in VRLEs is the evaluation of process data or learning analytics. These approaches might offer further insights into the underlying cognitive learning processes (Christopoulos, Pellas, & Laakso, 2020). In this way, the effects of the different support approaches could be explored in more detail. However, initial approaches in this direction point to a need for further research to assign the process data to the corresponding learning processes (Schumacher & Ifenthaler, 2021).

Depending on the chosen learning content and setting presented in VR, the problem of motion sickness might arise. In the present thesis, the included simulations did not impose a fast-changing environment. However, displaying fast-moving VR animation might induce motion sickness (Kim, Park, Choi, & Choe, 2018). This factor should therefore be carefully considered when selecting VR content, as systematic testing of motion sickness involves ethical challenges. In addition to motion sickness, learners might suffer from visual strain imposed by the stereoscopic presentation of learning content (Moreau, 2013). In the context of learning in VR, visual strain might decrease learning outcomes. Here, future research might answer the question of how long learning units in VR should last and how long breaks between two VR sessions should be.

Long-term effects of learning in virtual reality

A major barrier for deep learning in VR is the novelty of this learning setting and a lack of adequate learning behavior. However, when using VRLEs regularly, this influencing factor might decrease in importance while other factors might become more relevant. Repeated learning in VR might reduce the potential novelty effect and might have an impact on the motivating character of VRLEs (Howard & Lee, 2019). With increased familiarization, the use of the technology might be eased, but the learners might need additional activation or motivation. Thus, the long-term effects of using VRLEs as a common tool while studying must be addressed in future research, as this thesis does not provide insights into long-term effects.

Moreover, in the included studies, the learning outcome and the mental model development were measured directly after learning. Learners therefore had little time to consolidate the learning content and to connect it to their prior knowledge network within their long-term memory (Kühl, Navratil, & Münzer, 2018). Previous findings have outlined that consolidation of new information might need time (Arndt, Schueler, & Scheiter, 2015). Based on this, further, follow-up measurements of the learning outcome and mental model development provide the opportunity to investigate how sustainable learning in VR was. Incorporating VRLEs into a regular university curriculum would allow for follow-up studies and open promising opportunities to study VRLEs in a more realistic setting than in the present thesis.

When using VRLEs regularly, the support approaches might be extended to encompass special training aimed at deeply processing the learning content. For conventional learning material, coherence formation training is effective and can substantially increase learning outcomes (Kombartzky, Ploetzner, Schlag, & Metz, 2010; Seufert, 2019). This training could be adapted to learning in VRLEs with a special focus on learning with VR animation and

narrations. Promising learning strategies could be taught and trained to achieve beneficial long-term effects on learning in VR (Kühl, Scheiter, Gerjets, & Gemballa, 2011; Kombartzky et al., 2010). In addition to training, cognitive learning strategies such as metacognitive monitoring of learning in VR might be important (Hasselhorn & Gold, 2017).

Further technical, as well as organizational issues, might need to be overcome before implementing VRLEs regularly in the everyday university setting (Bricken, 1991). Particularly, group settings and collaborative settings need to be further explored (Hwang & Hu, 2013). Depending on the used devices, overlapping tracking spaces might cause interference in group settings. Furthermore, cooperative settings while learning in VR need to be explored and optimized in future studies.

Extending statistical analyses

Including VRLEs in the university context would allow for the use of different statistical approaches to analyze research findings as a result of larger sample sizes and available long-term effects. The included power analysis in this thesis was based on findings from conventional learning settings and the effect sizes were only briefly estimated. From the present findings, more precise a priori power analyses for the different support approaches in VR are now available, and thus, statistical power can be better managed in future studies. In general, higher sample sizes would allow for more complex analysis of the potential interplay of the different factors. With the addition of a larger sample size, a more complex multi-level analysis could be conducted including different mediators and moderators or mediated moderations or moderated mediations (Zitzmann & Helm, 2021). Additionally, when using long-term data, autoregression models such as crossed lagged panel models would allow for further insights into the development of expertise or motivation during the course (Krull, Wilbert, & Hennemann, 2018). Furthermore, collecting process data over a long time might shed further light on challenges while learning and possibilities to adapt the VRLEs (Christopoulos et al., 2020). Collecting more data might offer the opportunity to choose different statistical approaches such as Bayesian hypothesis-testing (Wagenmakers et al., 2018). One advantage of this approach is that prior findings can be included in the analysis to overcome the limitations of simply dichotomous testing of hypotheses based on p -values (Wagenmakers et al., 2018). Using the available data, different models can be contrasted with the central influencing factors and their differences can be quantified (Rouder, 2014).

5.2. Conclusion

The primary aim of the present thesis was to outline the potential of VRLEs as a contemporary learning environment for universities. Within this, the questions of what makes learning in VR special and what specific challenges learners might face while learning in VR were explored. By using the theoretical framework EMIL VR, different support approaches were systematically examined under the consideration of learner's characteristics. As the learning outcome was measured in a differentiated manner, and in one study, using a multi-method approach, the present research also shed light on the associated underlying cognitive learning processes as well as mental model development. All chosen support approaches increased learning outcomes in different ways and on different levels. To this end, even pragmatic and initially quite basic support approaches, such as the elaboration prompt, were found to have a significant positive effect on learning outcomes. In summary, the present findings imply that the key to success is to choose support approaches that are related to challenging subprocesses while learning in VR. Strategic activations, as well as appropriate design of the VRLEs, are pragmatic and effective ways to foster learning in VR. Adding to this, the importance of considering learner's characteristics was also emphasized by the present findings. In terms of future research, the theoretical framework, EMIL VR, could be extended and more global support approaches could be included to support learners in VR in the long term. Important factors such as cognitive or perceptual load need to be investigated. Combining VRLEs with further scientific advancements such as artificial intelligence might extend the benefits of learning in VR. Future research might also develop adaptive VRLEs that provide different versions of VR to support individual learners on their learning paths. This could open a new way to perceive, process, and integrate complex learning content and might thus reshape the learning culture at the university in a unique way.

References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. doi: 10.1016/j.learninstruc.2006.03.001
- Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, 166, 104154. doi: 10.1016/j.compedu.2021.104154
- Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development*, 68(5), 2095–2119. doi: 10.1007/s11423-020-09748-7
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. Longman.
- Arndt, J., Schüler, A., & Scheiter, K. (2019). Investigating the influence of simultaneous–versus sequential–text-picture presentation on text-picture integration. *The Journal of Experimental Education*, 87(1), 116–127. doi: 10.1080/00220973.2017.1363690
- Arndt, J., Schueler, A., & Scheiter, K. (2015). Text–picture integration: How delayed testing moderates recognition of pictorial information in multimedia learning. *Applied Cognitive Psychology*, 29(5), 702–712. doi: 10.1002/acp.3154
- Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 254–270). Routledge.
- Babu, S. K., Krishna, S., Unnikrishnan, R., & Bhavani, R. R. (2018). Virtual reality learning environments for vocational education: A comparison study with conventional instructional media on knowledge retention. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 385–389). doi: 10.1109/ICALT.2018.00094
- Baceviciute, S., Mottelson, A., Terkildsen, T., & Makransky, G. (2020). Investigating representation of text and audio in educational VR using learning outcomes and EEG. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. ACM. doi: 10.1145/3313831.3376872
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. doi: 10.1126/science.1736359
- Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. does the verbalisation method affect learning? *Metacognition and Learning*, 3(1), 39–58. doi: 10.1007/s11409-007-9009-

- Bannert, M., & Reimann, P. (2011). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. doi: 10.1007/s11251-011-9167-4
- Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short-and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, 52, 293–306. doi: 10.1016/j.chb.2015.05.038
- Bell, J. T., & Fogler, H. S. (1995). The investigation and application of virtual reality as an educational tool. In *Proceedings of the American Society for Engineering Education Annual Conference* (pp. 1718–1728).
- Berthold, K., Eysink, T. H. S., & Renkl, A. (2008). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, 37(4), 345–363. doi: 10.1007/s11251-008-9051-z
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? the role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577. doi: 10.1016/j.learninstruc.2007.09.007
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of educational psychology*, 101(1), 70. doi: 10.1037/a0013247
- Biggs, J. (1989). Approaches to the enhancement of tertiary teaching. *Higher Educational Research and Development*, 8(1), 7-25. doi: 10.1080/0729436890080102
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: a meta-analysis. *Educational Psychology Review*, 30(3), 703–725. doi: 10.1007/s10648-018-9434-x
- Bloom, B. (1956). *A taxonomy of cognitive objectives* (Handbook I: The cognitive domain. ed.). David McKay CoInc.
- Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, 30(7), 709–725. doi: 10.1080/09588221.2017.1356335
- Bricken, M. (1991). Virtual reality learning environments: potentials and challenges. *Acm Siggraph Computer Graphics*, 25(3), 178–184. doi: 10.1145/126640.126657
- Britton, B. K., Glynn, S. M., Meyer, B. J., & Penland, M. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74(1), 51.
- Brunyé, T. T., Taylor, H. A., & Rapp, D. N. (2008). Repetition and dual coding in procedural multimedia presentations. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(7), 877–895. doi: 10.1002/acp.1396
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98(1), 182–197. doi: 10.1037/0022-0663.98.1.182
- Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. *Computers & Education*,

- 159, 104005. doi: 10.1016/j.compedu.2020.104005
- Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15(3), 139–164. doi: 10.1111/j.1745-3984.1978.tb00065.x
- Chen, C. J., Toh, S. C., & Ismail, W. M. F. W. (2005). Are learning styles relevant to virtual reality? *Journal of Research on Technology in Education*, 38(2), 123–141. doi: 10.1080/15391523.2005.10782453
- Chen, Y. L. (2016). The effects of virtual reality learning environment on student cognitive and linguistic development. *The Asia-Pacific Education Researcher*, 25(4), 637–646. doi: 10.1007/s40299-016-0293-2
- Christopoulos, A., Pellas, N., & Laakso, M.-J. (2020). A learning analytics theoretical framework for stem education virtual reality applications. *Education Sciences*, 10(11), 317. doi: 10.3390/educsci10110317
- Clark, R. C., Nguyen, F., & Sweller, J. (2011). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of educational research*, 53(4), 445–459. doi: 10.3102/00346543053004445
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7(12), 547–552. doi: 10.1016/j.tics.2003.10.005
- Craik, K. (1943). *The nature of explanation*. Cambridge University Press.
- Daumiller, M., & Dresel, M. (2019). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education*, 87(1), 161–176. doi: 10.1080/00220973.2018.1448744
- Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: Validity and utility of structural knowledge assessment. *Journal of Business and Psychology*, 18(2), 191–206. doi: 10.1023/a:1027397031207
- de Bruin, A. B., Roelle, J., Carpenter, S. K., & Baars, M. (2020). Synthesizing cognitive load and self-regulation theory: a theoretical framework and research agenda. *Educational Psychology Review*, 32, 903–915. doi: 10.1007/s10648-020-09576-4
- de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, 21(2), 113–140. doi: 10.1007/s10648-009-9098-7
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, 21(6), 731–746. doi: 10.1002/acp.1346
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education*, 55(2), 681–691. doi: 10.1016/j.compedu.2010.02.027
- Dev, P. C. (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? *Remedial and special education*, 18(1), 12–19. doi:

- 10.1177/074193259701800104
- Díaz, P., & Ioannou, A. (2019). *Learning in a digital world: An introduction*. Singapore: Springer Nature.
- Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in k-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, 51(6), 2006–2033. doi: 10.1111/bjet.13030
- Dörrenbächer, L., & Perels, F. (2016). Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning. *Learning and Individual Differences*, 51, 229–241. doi: 10.1016/j.lindif.2016.09.015
- Eckert, A. (2000). Die Netzwerk-Elaborations-Technik (N E T) –Ein computerunterstützendes Verfahren zur Diagnose komplexer Wissensstrukturen. In H. M. . F. Fischer (Ed.), (pp. 137–157). Hogrefe.
- Eitel, A., & Scheiter, K. (2014, apr). Picture or text first? explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, 27(1), 153–180. doi: 10.1007/s10648-014-9264-4
- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13–20. doi: 10.1016/j.learninstruc.2016.11.010
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using GPower 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. doi: 10.3758/brm.41.4.1149
- Fiorella, L., & Mayer, R. E. (2015). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. doi: 10.1007/s10648-015-9348-9
- Folker, S., Ritter, H., & Sichelschmidt, L. (2005). Processing and integrating multimodal material—the influence of color-coding. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 27).
- Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, 33(2), 739–762. doi: 10.1007/s10648-020-09561-x
- Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7(2), 155–170. doi: 10.1207/s15516709cog0702_3
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52(1), 45. doi: 10.1037/0003-066X.52.1.45
- Gentner, D., Rattermann, M., & Forbus, K. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25(4), 524–575. doi: 10.1006/cogp.1993.1013
- Gentner, N., & Seufert, T. (2020). The double-edged interactions of prompts and self-efficacy. *Metacognition and Learning*, 15(2), 261–289. doi: 10.1007/s11409-020-09227-7
- Goodwin, M. S., Wiltshire, T., & Fiore, S. M. (2015). Applying research in the cognitive sciences to the design and delivery of instruction in virtual reality learning environments.

- In *International conference on virtual, augmented and mixed reality* (pp. 280–291). doi: 10.1007/978-3-319-21067-4_29
- Goodyear, P., & Retalis, S. (2010). *Technology-enhanced learning: Design patterns and pattern languages*. Brill.
- Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology*, 70(5), 657–669. doi: /10.1037/0022-0663.70.5.657
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 220–227. doi: 10.1177/2372732216655542
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, 17(3), 300 – 312. doi: 10.1037/0012-1649.17.3.300
- Hasselhorn, M., & Gold, A. (2017). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren*. Kohlhammer Verlag.
- Hedberg, J., & Alexander, S. (1994). Virtual reality in education: Defining researchable issues. *Educational Media International*, 31(4), 214–220. doi: 10.1080/0952398940310402
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in cognitive science*, 3(3), 446–474. doi: 10.1111/j.1756-8765.2011.01150.x
- Höfler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations. evidence for an ability as compensator hypothesis. *Computers in Human Behavior*, 27(1), 209–216. doi: 10.1016/j.chb.2010.07.042
- Howard, M. C., & Lee, J. (2019). Pre-training interventions to counteract seductive details in virtual reality training programs. *Human Resource Development Quarterly*, 31(1), 13–29. doi: 10.1002/hrdq.21378
- Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A.-S. (2020). Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computing Research*, 58(3), 596–615. doi: 10.1177/0735633119867422
- Hu-Au, E., & Okita, S. (2021). Exploring differences in student learning and behavior between real-life and virtual reality chemistry laboratories. *Journal of Science Education and Technology*, 1–15. doi: 10.1007/s10956-021-09925-0
- Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning*, 22(6), 392–404. doi: 10.1111/j.1365-2729.2006.00180.x
- Hwang, W.-Y., & Hu, S.-S. (2013). Analysis of peer learning behaviors using multiple representations in virtual reality and their impacts on geometry problem solving. *Computers & Education*, 62, 308–319. doi: 10.1016/j.compedu.2012.10.005
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society*, 15(1), 38–52.
- Jenkins, J. J. (1979). Four points to remember: a tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory* (p. 429–446).

- Hillsdale, NJ: Erlbaum.
- Johnson, A. M., Reisslein, J., & Reisslein, M. (2014). Representation sequencing in computer-based engineering education. *Computers & Education*, 72, 249–261. doi: 10.1016/j.compedu.2013.11.010
- Johnson-Glenberg, M. C. (2019). The necessary nine: Design principles for embodied vr and active stem education. In *Learning in a Digital World* (pp. 83–112). Springer. doi: 10.1007/978-981-13-8265-9-5
- Johnson-Laird, P. N. (1981). Mental models in cognitive science. *Cognitive Science*, 4(1), 71–115. doi: 10.1016/s0364-0213(81)80005-5
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.
- Johnson-Laird, P. N. (1996). *Images, models, and propositional representations*.
- Jonassen, D. H., & Henning, P. (1999). Mental models: Knowledge in the head and knowledge in the world. *Educational technology*, 37–42.
- Kalyuga, S. (2009). The expertise reversal effect. In *Managing Cognitive Load in Adaptive Multimedia Learning*. IGI Global. doi: 10.4018/978-1-60566-048-6
- Kamarainen, A. M., Thompson, M., Metcalf, S. J., Grotzer, T. A., Tutwiler, M. S., & Dede, C. (2018). Prompting connections between content and context: Blending immersive virtual environments and augmented reality for environmental science learning. In *Communications in computer and information science* (pp. 36–54). Springer International Publishing. doi: 10.1007/978-3-319-93596-6_3
- Kardong-Edgren, S. S., Farra, S. L., Alinier, G., & Young, H. M. (2019). A call to unify definitions of virtual reality. *Clinical Simulation in Nursing*, 31, 28–34. doi: 10.1016/j.ecns.2019.02.006
- Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69, 66–73. doi: 10.1016/j.apergo.2017.12.016
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.01997
- Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology*, 51(6), 2115–2138. doi: 10.1111/bjet.13029
- Kollöffel, B. (2012). Exploring the relation between visualizer–verbalizer cognitive styles and performance with visual or verbal learning material. *Computers & Education*, 58(2), 697–706. doi: 10.1016/j.compedu.2011.09.016
- Kombartzky, U., Ploetzner, R., Schlag, S., & Metz, B. (2010). Developing and evaluating a strategy for learning from animations. *Learning and Instruction*, 20(5), 424–433. doi: 10.1016/j.learninstruc.2009.05.002
- 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(2), 311–328. doi: 10.1037/0021-9010.78.2.311
- Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European Journal of Psychology of Education*, 14(1), 23–40. doi: 10.1007/BF03173109
- Krull, J., Wilbert, J., & Hennemann, T. (2018). Does social exclusion by classmates lead to behaviour problems and learning difficulties or vice versa? A cross-lagged panel analysis. *European Journal of Special Needs Education*, 33(2), 235–253. doi: 10.1080/08856257.2018.1424780
- Kühl, T., Navratil, S. D., & Münzer, S. (2018). Animations and static pictures: The influence of prompting and time of testing. *Learning and Instruction*, 58, 201–209. doi: 10.1016/j.learninstruc.2018.07.006
- Kühl, T., Scheiter, K., Gerjets, P., & Gemballa, S. (2011). Can differences in learning strategies explain the benefits of learning from static and dynamic visualizations? *Computers & Education*, 56(1), 176–187. doi: 10.1016/j.compedu.2010.08.008
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100. doi: 10.1016/S0364-0213(87)80026-5
- Lee, E. A.-L., & Wong, K. W. (2008). A review of using virtual reality for learning. *Transactions on Edutainment*, 231–241. doi: 10.1007/978-3-540-69744-2_18
- Lee, E. A.-L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, 79, 49–58. doi: 10.1016/j.compedu.2014.07.010
- Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, 97(2), 184–196. doi: 10.1037/0022-0663.97.2.184
- Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, 29, 174–175. doi: 10.1016/j.learninstruc.2013.05.004
- Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction*, 13(2), 157–176. doi: 10.1016/S0959-4752(02)00018-X
- Makransky, G., Andreassen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, 4(113), 719–735. doi: 10.1037/edu0000473
- Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6), 691–707. doi: 10.1111/jcal.12375
- Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, 72, 276–285. doi: 10.1016/j.chb.2017.02.066
- Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality.

- Educational Psychology Review*, 1–22. doi: 10.1007/s10648-020-09586-2
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. doi: 10.1016/j.learninstruc.2017.12.007
- Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly*, 48(2), 199–214. doi: 10.1002/rrq.41
- Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, 47(6), 1083–1095. doi: 10.1111/bjet.12271
- Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 93(2), 377. doi: 10.1037/0022-0663.93.2.377
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American psychologist*, 59(1), 14–19. doi: 10.1037/0003-066X.59.1.14
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41, 31–48.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171. doi: 10.1016/j.learninstruc.2009.02.012
- Mayer, R. E. (Ed.). (2014). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press. doi: 10.1017/cbo9781139547369
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. doi: 10.1111/jcal.12197
- Mayer, R. E. (2020). Immersion principle. In R. Mayer (Ed.), *Multimedia Learning - Third Edition* (p. 375–369). Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484. doi: 10.1037/0022-0663.83.4.484
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88(1), 64. doi: 10.1037/0022-0663.88.1.64
- Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4), 715–726. doi: 10.1037/0022-0663.82.4.715
- Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, 8(3), 147–154. doi: 10.1037/1076-898x.8.3.147
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389–401. doi: 10.1037/0022-0663.86.3.389
- Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook

- design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, 43(1), 31–41. doi: 10.1007/bf02300480
- Mazuryk, T., & Gervautz, M. (1996). Virtual reality-history, applications, technology and future. *Virtual Reality*.
- McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology*, 34(2), 143–153. doi: 10.1016/j.cedpsych.2008.12.003
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40. doi: 10.1016/j.compedu.2013.07.033
- Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140, 103603. doi: 10.1016/j.compedu.2019.103603
- Miller, M. D., Castillo, G., Medoff, N., & Hardy, A. (2021). Immersive VR for organic chemistry impacts on performance and grades for first-generation and continuing-generation university students. *Innovative Higher Education*, 1–25. doi: 10.1007/s10755-021-09551-z
- Miranda, A. T., & Palmer, E. M. (2013). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods*, 46(1), 159–172. doi: 10.3758/s13428-013-0357-7
- Moreau, G. (2013). Visual immersion issues in Virtual Reality: A survey. In *26th Conference on Graphics, Patterns and Images Tutorials* (pp. 6–14). doi: 10.1109/SIBGRAPI-T.2013.9
- Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of educational Psychology*, 96(1), 165. doi: 10.1037/0022-0663.96.1.165
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. doi: 10.1007/s10648-007-9047-2
- Mulders, M., Buchner, J., & Kerres, M. (2020). A framework for the use of immersive virtual reality in learning environments. *International Journal of Emerging Technologies in Learning (iJET)*, 15(24), 208–224.
- Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352–1375. doi: 10.1080/00140130802170387
- Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, 19(3), 259–271. doi: 10.1016/j.learninstruc.2008.05.002
- Onyesolu, M. O., & Eze, F. U. (2011). Understanding virtual reality technology: advances and applications. *Advances in Computer Science and Engineering*, 53–70.
- Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? evidence from eye movements. *Computers in Human Behavior*, 26(1), 110–117. doi: 10.1016/j.chb.2009.09.001
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University

- Press.
- Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395–417.
- Park, B., Münzer, S., Seufert, T., & Brünken, R. (2016). The role of spatial ability when fostering mental animation in multimedia learning: An ati-study. *Computers in Human Behavior*, 64, 497–506. doi: 10.1016/j.chb.2016.07.022
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785. doi: 10.1037/edu0000241
- Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, 37(1), 226–241. doi: 10.1111/jcal.12482
- Pieger, E., & Bannert, M. (2018). Differential effects of students’ self-directed metacognitive prompts. *Computers in Human Behavior*, 86, 165–173. doi: 10.1016/j.chb.2018.04.022
- Ploetzner, R., & Lowe, R. (2012). A systematic characterisation of expository animations. *Computers in Human Behavior*, 28(3), 781–794. doi: 10.1016/j.chb.2011.12.001
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12(1), 61–86. doi: 10.1016/S0959-4752(01)00016-0
- Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., & Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. *Computers & Education*, 95, 309–327. doi: 10.1016/j.compedu.2016.02.002
- Prouzeau, A., Wang, Y., Ens, B., Willett, W., & Dwyer, T. (2020). Corsican twin: Authoring in situ augmented reality visualisations in virtual reality. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 1–9).
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. doi: 10.1016/j.compedu.2019.103778
- Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7), 794–807.
- Rapp, D. N. (2005). Mental models: Theoretical issues for visualizations in science education. In *Visualization in Science Education* (pp. 43–60). Springer. doi: 10.1007/1-4020-3613-2_4
- Reigeluth, C., & Stein, R. (1983). Elaboration theory. *Instructional-design theories and models: An overview of their current status (1983)*, 335–381.
- Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 229–247). Cambridge University Press.
- Rheinberg, F. (2006). Intrinsische Motivation und Flow-Erleben. In *Motivation und Handeln* (pp. 331–354). Springer.

- Richter, J., & Scheiter, K. (2019). Studying the expertise reversal of the multimedia signaling effect at a process level: Evidence from eye tracking. *Instructional Science*, 47(6), 627–658. doi: 10.1007/s11251-019-09492-3
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. doi: 10.1016/j.edurev.2015.12.003
- Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text–picture relations in multimedia learning: The influence of prior knowledge. *Journal of Educational Psychology*, 110(4), 544–560. doi: 10.1037/edu0000220
- Riva, G., Davide, F., & IJsselstein, W. A. (2003). *Being there: Concepts, effects and measurements of user presence in synthetic environments*. Ios Press.
- Roediger, H. L., III. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, 59, 225–254. doi: 10.1146/annurev.psych.57.102904.190139
- Roelle, J., & Berthold, K. (2013). The expertise reversal effect in prompting focused processing of instructional explanations. *Instructional Science*, 41(4), 635–656. doi: 10.1007/s11251-012-9247-0
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2), 301–308. doi: 10.3758/s13423-014-0595-4
- Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349–363. doi: 10.1037/0033-2909.100.3.349
- Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. doi: 10.1006/ceps.1999.1020
- Ryan, R. M., & Deci, E. L. (2000b). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In *Intrinsic and Extrinsic Motivation* (pp. 13–54). Elsevier. doi: 10.1016/B978-012619070-0/50024-6
- Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Hogrefe Göttingen.
- Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence*, 46, 156–168. doi: 10.1016/j.intell.2014.05.014
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. doi: 10.1016/s0959-4752(02)00017-8
- Schroeder, S. (2011). What readers have and do: Effects of students’ verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103(4), 877. doi: 10.1037/a0023731
- Schumacher, C., & Ifenthaler, D. (2021). Investigating prompts for supporting students’ self-regulation—a remaining challenge for learning analytics approaches? *The Internet and Higher Education*, 49, 100791. doi: 10.1016/j.iheduc.2020.100791
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations.

- Learning and Instruction*, 13(2), 227–237. doi: 10.1016/s0959-4752(02)00022-1
- Seufert, T. (2019, feb). Training for coherence formation when learning from text and picture and the interplay with learners' prior knowledge. *Frontiers in Psychology*, 10. doi: 10.3389/fpsyg.2019.00193
- Simon, H. A. (1978). *On the forms of mental representation*. University of Minnesota Press, Minneapolis.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431–433. doi: 10.1111/bjop.12305
- Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13–59). New York: Freeman.
- Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (pp. 37–76). Elsevier. doi: 10.1016/b978-0-12-387691-1.00002-8
- Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior*, 39, 356–367. doi: 10.1016/j.chb.2014.07.018
- Thillmann, H., Kunsting, J., Wirth, J., & Leutner, D. (2009). Is it merely a question of “what” to prompt or also “when” to prompt? the role of point of presentation time of prompts in self-regulated learning. *Zeitschrift für Pädagogische Psychologie*, 23(2), 105–115. doi: 10.1024/1010-0652.23.2.105
- Thorsteinsson, G., & Page, T. (2008). Reflecting from the literature on constructivist learning in virtual reality in relation to the FISTE comenius 2.1 project. *Educacia*, 21, 95–104.
- Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64(1), 37–54. doi: 10.3102/00346543064001037
- Trumpower, D. L., & Vanapalli, A. S. (2016). Structural assessment of knowledge as, of, and for learning. In *Learning, Design, and Technology* (pp. 1–22). Springer International Publishing. doi: 10.1007/978-3-319-17727-4_23-1
- Tversky, B., Heiser, J., Mackenzie, R., Lozano, S., & Morrison, J. (2008). Learning with Animation: Research Implications for Design. In R. Lowe & W. Schnotz (Eds.), (pp. 263–285). Cambridge University Press New York.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International journal of human-computer studies*, 57(4), 247–262. doi: 10.1006/ijhc.2002.1017
- Ullmann, S. (1984). Visual routines. *Cognition*, 18, 97–159.
- van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction*, 16(3), 199–212. doi: 10.1016/j.learninstruc.2006.03.007
- van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. Mayer (Ed.), *The cambridge handbook of multimedia learning* (pp. 263–278). Cambridge University Press. doi: 10.1017/cbo9781139547369.014
- Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent

- developments and future directions. *Educational Psychology Review*, 17(2), 147–177. doi: 10.1007/s10648-005-3951-0
- Vogt, A., Albus, P., Montag, M., Drey, T., Hartwig, S., Tasliarmut, F., & Wolf, M. (2021). *Effekte von Segmentierung und Color Coding in nicht-immersiven Virtual-Reality-Lernumgebungen*. 18. Fachgruppentagung der Fachgruppe Pädagogische Psychologie (PAEPSY): Heidelberg/online, Germany.
- Vogt, A., Albus, P., & Seufert, T. (2021). Learning in virtual reality: Bridging the motivation gap by adding annotations. *Frontiers in Psychology*, 12, 881. doi: 10.3389/fpsyg.2021.645032
- Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021a). Immersive virtual reality or auditory text first? effects of adequate sequencing and prompting on learning outcome. *British Journal of Educational Technology*. doi: 10.1111/bjet.13104
- Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021b). Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model. *Computers & Education*, 104235. doi: 10.1016/j.compedu.2021.104235
- Vogt, A., Klepsch, M., Baetge, I., & Seufert, T. (2020). Learning from multiple representations: Prior knowledge moderates the beneficial effects of signals and abstract graphics. *Frontiers in Psychology*, 11, 3462. doi: 10.3389/fpsyg.2020.601125
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... others (2018). Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, 25(1), 35–57. doi: 10.3758/s13423-017-1343-3
- Wallen, E., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development*, 53(3), 59–71. doi: 10.1007/BF02504798
- Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24(6), 1319–1333. doi: 10.1080/10494820.2014.994533
- Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. *New directions for teaching and learning*, 2011(126), 45–53. doi: 10.1002/tl.443
- White, B. Y., & Frederiksen, J. R. (1986). Progressions of qualitative models as a foundation for intelligent learning environments. *Report No. 6277*.
- Winn, W. (1982). The role of diagrammatic representation in learning sequences, identification and classification as a function of verbal and spatial ability. *Journal of Research in Science Teaching*, 19(1), 79–89. doi: 10.1002/tea.3660190110
- Wischgoll, A. (2016). Combined training of one cognitive and one metacognitive strategy improves academic writing skills. *Frontiers in Psychology*, 7, 187. doi: 10.3389/fpsyg.2016.00187
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005. doi: 10.1111/bjet.13023
- Zitzmann, S., & Helm, C. (2021). Multilevel analysis of mediation, moderation, and

References

nonlinear effects in small samples, using expected a posteriori estimates of factor scores. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–18. doi: 10.1080/10705511.2020.1855076

Part IV.

Appendix

Appendix A - Peer-reviewed journal articles

Appendix A - Peer-reviewed journal articles

The first paper was published in *Frontiers in Psychology* under the terms of the Creative Commons CC-BY license (<https://creativecommons.org/licenses/by/4.0/>):

Vogt, A., Albus, P., & Seufert, T. (2021). Learning in virtual reality: Bridging the motivation gap by adding annotations. *Frontiers in Psychology*, 12, 645032.
<https://doi.org/10.3389/fpsyg.2021.645032>

Journal Metrics: 2-Year Impact Factor: 2.990 (2020), h-index 110



Learning in Virtual Reality: Bridging the Motivation Gap by Adding Annotations

Andrea Vogt*, Patrick Albus and Tina Seufert

Department Learning and Instruction, Institute of Psychology and Education, Ulm University, Ulm, Germany

OPEN ACCESS

Edited by:

Bernhard Ertl,
Munich University of the Federal
Armed Forces, Germany

Reviewed by:

Kenneth Lim,
Nanyang Technological
University, Singapore
Lijia Lin,
East China Normal University, China

*Correspondence:

Andrea Vogt
andrea.vogt@uni-ulm.de

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 22 December 2020

Accepted: 01 March 2021

Published: 24 March 2021

Citation:

Vogt A, Albus P and Seufert T (2021)
Learning in Virtual Reality: Bridging the
Motivation Gap by Adding
Annotations.
Front. Psychol. 12:645032.
doi: 10.3389/fpsyg.2021.645032

One challenge while learning scientific concepts is to select relevant information and to integrate different representations of the learning content into one coherent mental model. Virtual reality learning environments (VRLEs) offer new possibilities to support learners and foster learning processes. Whether learning in VR is successful, however, depends to a large extent on the design of the VRLE and the learners themselves. Hence, adding supportive elements in VRLEs, such as annotations, might facilitate the learning process by guiding attention and supporting the selection of relevant information. Additionally, the mapping of pictorial and verbal information is eased by these annotations. The beneficial effect of annotations is highly dependent on learners' intrinsic motivation as intrinsic motivation while learning also affects the information selection and visual search patterns. In our experimental study ($N = 61$), we compared two conditions: learning in a VRLE with or without annotations. We measured the learning outcome on three different levels (knowledge, comprehension, and application). Additionally, we investigated intrinsic motivation as a moderator for the effect of annotations on learning outcome. We found no significant main effect of annotations on learning outcome. The moderating effect of intrinsic motivation for annotations on the overall learning outcome was significant. Our results imply that learners are either intrinsically motivated or need additional support by annotations as these support the selection of relevant information in the VRLE and therefore enable them to learn successfully. Which type or quantity of annotations supports learning processes best needs to be explored in future research.

Keywords: virtual reality, signaling, intrinsic motivation, multiple representations in multimedia learning, coherence formation, mental models

INTRODUCTION

Many different approaches to designing learning environments and supporting learners exist. Particularly, for science, often abstract representations of the learning content (e.g., in textbooks) challenges the learners to process, translate, and include concepts into their mental model (Rapp, 2007). As learning scientific concepts is *per se* challenging, the learners need to be supported by an appropriate presentation of the learning content. Demonstrations are such an appropriate form, which allows learners to explore scientific processes, such as chemical reactions. However, demonstrations cannot be used in all cases. Some processes might occur on a micro level and may not be observable (Zacharia, 2007; Akçayir et al., 2016). Some demonstrations are not applicable to a university setting because they need a complex technical setup, are too dangerous, or too

expensive (e.g., Wilkerson-Jerde et al., 2015). In this case, new technologies, such as virtual reality (VR), can open the way to advanced education in different scientific domains (Potkonjak et al., 2016; Wu et al., 2020).

Using VR or 3D simulations to design learning environments offers the possibility to display a contextualized setting, but additionally provide the opportunity to analyze learning content in a different way (Fowler, 2015; Parong and Mayer, 2018; Radianti et al., 2020). These learning environments offer great opportunities for experimental learning and are described to be motivating, increase liking and engagement compared to conventional settings (Makransky et al., 2019; Di Natale et al., 2020; Klingenberg et al., 2020; Radianti et al., 2020). For example, invisible content can be made visible, and sensitive structures can be composed and decomposed flexibly. Furthermore, micro or macro phenomena, such as chemical reactions can be made visible in virtual reality learning environments (VRLEs) that cannot be observed in the physical world (Schott and Marshall, 2018; Han, 2020; Wu et al., 2020). Furthermore, potential risks can be overcome (Radianti et al., 2020). In contrast to virtual 3D animations, learners can immerse themselves in the VRLE. By immersion and the novelty effect of the VRLEs, learning processes might be additionally stimulated (Wu et al., 2020).

One challenge remains while learning science in a VRLE: Learners are usually confronted with different, multiple representations, such as text and pictures. The stereoscopic view of the VRLE impacts their perception and therefore might influence their cognitive learning processes (e.g., Paes et al., 2017). Mostly, next to the pre-dominant visual, pictorial presentation of the learning content, often auditory texts or short verbal labels or annotations are included (Mikropoulos and Natsis, 2011). All these different representations have to be understood and integrated into a coherent whole. For this integrative process of coherence formation, it is necessary to process the learning content deeply to identify and understand the relevant concepts and to integrate them (i.e., Bloom, 1956; Schnotz and Bannert, 2003). A common approach to facilitating such a deep-learning is to include highlights or signals to indicate relevant components of the learning content (Richter et al., 2016). Furthermore, the learners can be supported in finding corresponding elements (Gentner et al., 1993). For instance, adding annotations into the pictorial representation might help to connect corresponding information and highlight important aspects of the learning content (McTigue, 2009). Therefore, annotations might support and activate the learners, induce deeper cognitive learning processes, and foster coherence formation (Mayer et al., 1995).

Based on instructional design research, we know that these effects depend on the learner and their aptitudes (Seufert, 2003). Adding, for example, annotations might support learners with lower prior knowledge, and might cause interference for learners with high prior knowledge (Kalyuga, 2009). Hence, learners' aptitude and state need to be considered. Factors such as intrinsic motivation might also have a substantial impact on invested resources, as well as on the cognitive processing of the learning material (Moreno and Mayer, 2007; Baranes et al., 2014; Eseryel et al., 2014; Miranda and Palmer, 2014).

Motivated learners show efficient visual search patterns which help them to learn successfully. Hence, annotations support learners with low motivation as they induce a more effective visual search pattern and consequently learning outcome is expected to be increased (Eseryel et al., 2014; Miranda and Palmer, 2014).

Recent studies compared VRLEs to other learning settings (e.g., Parong and Mayer, 2018; Makransky et al., 2019) and corroborate that VRLEs could be used for learning. However, these previous approaches and publications have not considered how VRLEs should be designed in order to support learning and the necessary cognitive processes. Most notably, they have not taken into account for whom VRLE should be designed in which way, particularly depending on learners' motivational state. Overall, the question of how a VRLEs needs to be designed to foster deep processing of the learning content, and how to support coherence formation depending on learners' motivational state, has not been answered yet. Therefore, the aim of the present study was to investigate the effect of annotations and their interaction with intrinsic motivation in VRLE on learning.

THEORETICAL BACKGROUND

The beneficial effects of annotations, the learners' intrinsic motivation, and the interplay of these factors will be addressed. To answer the question of how VRLEs should be designed for complex scientific content to foster learning and coherence formation, it is worth having a closer look at the relevant models that describe cognitive processes underlying learning with multiple representations.

Cognitive Processes While Learning With Multiple Representations in Virtual Reality

One theory to describe learning processes is the *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2014). One assumption of this theory is that information is processed via two different channels (auditory and visual). It states that information is coded in two memory systems: the visual and the verbal system. This idea is based on the *dual-code theory* proposed by Paivio (1990). Thus, it is assumed that two separate mental models are built from these two information sources. At the end of the learning process, these two different mental models are integrated into one coherent mental model. This happens under consideration of the individual prior knowledge stored in the long-term memory (Mayer, 2014). Another assumption is that working memory only has a limited capacity to process information. This goes along with the finding that working memory overload can occur when too much information has to be processed at the same time (Baddeley, 1992). Furthermore, the CTML assumes that processing information is an active process, including selecting, organizing, and integrating information. This reflects the active role of the learner and emphasizes that the learner has to invest certain resources to learn successfully.

To better understand the integration process of verbal and pictorial information into the mental model, the *Integrated Model*

of *Text and Picture Comprehension* (IMTPC) delivers more insights (Schnotz and Bannert, 2003). The *descriptive branch* describes the processing of verbal and textual information, and the *depictive branch* describes the processing of pictures.

Verbal information is first processed sub-semantically, which results in an internal representation based on the text surface. This enables learners to simply process the information without a deeper understanding and allows them to recall simple definitions or facts. Based on this superficially processed information further semantic processing takes place and results in a propositional representation or network. At this level, a learner would be able to understand the concepts. If learners need to mentally operate on the information and to apply it in other contexts, this propositional representation has to be translated into an analog representation, the mental model (Schnotz and Bannert, 2003). With this last step, learners switch from the descriptive to the depictive branch.

In contrast to this multi-level processing of textual representations, pictures are processed more directly *via* the depictive branch hence this branch is particularly important for learning in VRLEs. First, by perceptually processing a picture, the information is selected based on cognitive schemata. To semantically process this information, a picture comprehension process takes place by mapping visuospatial and semantic relations (Gentner et al., 1993; Schnotz and Bannert, 2003). Unlike the information processing in the descriptive branch, no additional translation of the processed information is needed to integrate the information because the created internal mental representation already has the same analog structure as the mental model (Schnotz and Bannert, 2003). Learners can also “read” and extract propositions from this mental model (i.e., they can again switch branches), in this case from the depictive to the descriptive side. Thus, there is an interplay between the descriptive and the depictive branch, and text and picture information can be integrated and mapped onto each other. However, neither the IMTPC nor the CTML describes in detail how this mapping and integration process takes place. Based on theories on learning with analogies or learning with multiple representations, these *mapping processes* can be distinguished into element-to-element and relation-to-relation mapping processes (Gentner et al., 1993; Seufert et al., 2007). Whereas element-to-element mapping processes enable the learner to connect the learning content on a superficial and syntactic level, relation-to-relation mapping processes refer to finding similarities on the semantic level (Seufert et al., 2007). Therefore, relational mapping provides support for developing a global understanding of the content, whereas element-based mapping supports connecting different components on a rather superficial level.

To deduce the different levels of processing, Bloom’s (1956) taxonomy is often used to measure learning outcome in a differentiated manner. Questions aiming at the first level, the *knowledge* level, refer to the recognition or recall of facts, terms, or basic concepts without necessarily understanding what they mean. Hence, this reflects the superficial and sub-semantic processing of learning content. The next level, the *comprehension*

level, involves semantic processing of the learning content. It refers to concepts or relational structures of single facts and can be derived when learners are for example, organizing them, describing them in summary form, or explaining the main ideas in their own words. For example, local relationships between the individual components can already be derived or concepts can be contrasted at this level. The next level of Bloom’s taxonomy already assumes that a correct and coherent mental model has been formed. At the *application* level, learners can for example, further decompose the learning content into components and determine how different parts interact with each other. Furthermore, learned basic principles can be transferred to other use cases.

Based on these models, and particularly based on the assumption of IMTPC that pictures can be processed more directly, VRLEs seem to be a very promising approach to displaying complex scientific content. VRLEs are visual worlds consisting of mostly pictorial representations, thus, they support learners more directly in building a coherent mental model. Even content that is invisible in the real world can be drafted in the VRLE and can be used as a scaffold for constructing a mental model. When a brief scaffold based on a pictorial representation is constructed, adding further details to the mental model through the descriptive branch is a more simple and efficient way to construct a coherent mental model than vice versa (Schüler et al., 2015). The IMTPC implies that the process of building a coherent mental model out of textual and pictorial information is not a simple one, and consumes cognitive resources of the learners (Schnotz and Bannert, 2003). Therefore, the question arises as to how learners can be supported by the design of the learning environment to foster coherent formation processes, especially in VRLEs.

Fostering Learning With Multiple Representations in Virtual Reality

In VR, the content is mainly displayed visually and the learners are confronted with highly salient visual or pictorial information. This perceptually demanding setting might lead to just-in-time processing, and learners might not be able to select and organize relevant information into a coherent mental model (Renkl and Scheiter, 2017). Therefore, an adequate design of the VRLE is one crucial precondition for learning success.

Two different approaches to support learners by the chosen design could be used: One approach would be to guide learners’ attention to relevant aspects in the pictorial representation by displaying visual cues. Cueing refers to guiding attention by non-content means such as coloring or arrows (e.g., Mautone and Mayer, 2001; De Koning et al., 2007). However, this approach would guide learners’ attention but would not provide additional support to map the auditory text to the given animation. Another approach is to use textual annotations to support both, selecting and integrating relevant information (Mayer et al., 1995). For example, the positive effect of annotating pictorial representations in learning scientific content was described by McTigue (2009). Learners with the additional textual annotation that labeled important components in the pictorial

representation had significantly higher comprehension scores compared to controls without such annotations. Moreover, Mayer and Gallini (1990) found positive effects of annotations if these referred to parts and steps of scientific concepts. Considering their function, inserted annotations could be an effective mean to support learning processes in VRLEs. The three functions of annotations in the present study derive from the following facts: Annotations are salient, provided just-in-time, exactly at the place where they are needed, and repeat relevant information.

First, annotations function as signal as they are salient and appear just-in-time. Based on their salience, Mayer et al. (1995) states that “*annotated illustrations can serve as a signal that help readers select relevant words and images*” (p.40). Such signals guide attention toward certain aspects of the learning content and reduce unnecessary visual search and thus improve learning outcome (Richter et al., 2016). Hence, when learning in VRLEs, annotations can be used to emphasize certain components of the VR animation. The effect of signaling in this environment could be additionally strengthened when these annotations appear simultaneously to the corresponding narration, just-in time when learners should listen to the auditory text and combine this with the corresponding visual entity in the VR animation.

The second function is that annotations support mapping and integration processes, as they do not only appear just-in-time but also just-in-place (Gentner et al., 1993; Mayer, 2014). Based on this spatial contiguity, annotations signalize which aspects need to be mapped and integrated (Mayer and Fiorella, 2014). Learners’ attention is drawn to the auditory text, the corresponding visual entity and a verbal label for this entity. Thus, the mapping process is eased as learners do not need to invest cognitive resources to find the corresponding elements (Mayer et al., 1995).

The third function is repetition. The textual annotation repeats crucial terms of the auditory text and is also redundant to the pictorial entity of the VR animation. Thus, based on the redundancy principle one might assume that such a repetition is irrelevant and learners could be distracted or extraneously overloaded (Kalyuga and Sweller, 2014; Mayer and Fiorella, 2014). However, as annotations usually do not display the same words as the narration but were only short labels, they can function as coherent information which adds to understanding (coherence principle; Mayer, 2005). Additionally, it is described that particularly learners with low prior knowledge benefit from redundant information (Adesope and Nesbit, 2012). This repetition of important aspects leads to higher learning outcome on the knowledge level (Paivio, 1990).

To sum up, the positive effect of short textual annotations on the learning process has already been explored in classical multimedia settings. Prior findings indicate that performance improvement can be found for recall of conceptual information [see also Boers et al. (2017)] and on higher levels of learning outcome, such as transfer or comprehension tests (McTigue, 2009; Mason et al., 2013). However, so far there have been no empirical studies examining the effects of annotations in a VRLE.

Despite the fact that the design of learning environments might support learners to process the information more

effectively, there is one additional factor that might moderate these supporting effects, namely the learner him- or herself.

Individual Requirements for Learning Success in a Virtual Reality Learning Environment

Aptitude-treatment interaction describes that while designing treatments for learners, one needs to consider their aptitudes. One important aptitude is prior knowledge. Depending on learners’ prior knowledge, their performance is either enhanced by additional help or they are able to compensate for requirements of the given learning setting without additional help (e.g., Seufert, 2003). Therefore, learners’ prior knowledge has often been taken into account when analyzing the impact of supporting conditions. However, as the effects of prior knowledge are well-researched, we want to address motivation as an additional important factor of the learner, which might be especially relevant in VRLEs. Based on the assumptions of the CTML, further development of the model reflects the importance of affective processes of learning: The *Cognitive Affective Theory of Learning with Media* (CATLM; Moreno and Mayer, 2007) postulates that cognitive processes are influenced by affective states, for instance by motivation. Hence, learning performance might be moderated by motivational states.

Researchers agree that learners who are motivated are more likely to persist on their task, and are more willing to engage and to expend effort for the required task completion (Di Serio et al., 2013). Some previous findings have suggested that using VRLE fosters learners’ intrinsic motivation when these materials are compared to conventional learning materials, such as textbooks (Wu et al., 2020). Intrinsic motivation refers to doing something (e.g., learning) because it is inherently interesting, or the learner perceives the process as enjoyable. It is described as fostering learning and results in high-quality learning, as well as fostering creative processes (Ryan and Deci, 2000). Even for short learning tasks, intrinsic motivation is crucial for learning successfully (e.g., Fransson, 1977).

In a game-based virtual learning environment, Eseryel et al. (2014) uncovered an interaction effect of design and intrinsic motivation on learning by complex problem-solving. To explain this complex interaction of cognitive and motivational factors, underlying visual information processing needs to be considered. Previous studies imply that a higher intrinsic motivation goes along with a more efficient visual search pattern of the given task (Baranes et al., 2014; Miranda and Palmer, 2014). Hence, selection of relevant information is more efficient. As VRLEs are highly visually demanding, efficient visual search patterns are crucial for learning success. Therefore, particularly intrinsic motivation might have an impact on early information processes while learning and may have an impact on further cognitive processing of the learning content. Thus, being intrinsically motivated might have the same supportive effect on learning as annotations, namely an improved or eased visual search and a deepened learning process. Hence, the question arises as to whether both approaches—annotations as a design feature

and motivation as a learner feature—might compensate for each other.

Present Study

We aimed to gain further insights into how to design VRLEs for learning scientific concepts by using multiple representations appropriately. We investigated whether adding textual annotations has a positive effect on different levels of learning outcome (knowledge, comprehension, and application) while considering learners' aptitudes. Annotations, due to their two supportive functions (signaling and mapping), aimed to facilitate the construction of a coherent mental model.

Based on our postulated assumptions, we raise the following research questions:

(Q1) *Do annotations have a positive impact on learning outcome?* Previous studies showed beneficial effects of adding annotations on different levels of learning outcome (e.g., Mayer and Gallini, 1990; McTigue, 2009; Mason et al., 2013). Due to their signaling nature, annotations are expected to support the learner on a superficial level because they facilitate visual search and guide attention (McTigue, 2009; Ozcelik et al., 2010). Additionally, the annotations used in the present study are assumed to facilitate element-to-element mapping processes because they did not include additional information about processes or relations of the learning content. They simply labeled the visually displayed learning content with one or two words. Therefore, the used annotations simply repeated certain aspects and highlight connections between the textual label and the visual entity (Seufert et al., 2007). Both functions of annotations are expected to have a significant impact on learning outcome on the knowledge level (Mayer and Gallini, 1990; Boers et al., 2017). Based on the taxonomy of Bloom (1956), the different levels of learning outcome are characterized as hierarchical. Hence, the chosen annotations might also increase the learning outcome on a higher level, but not as much as on the knowledge level (Broadbent, 2017). Based on these discussed underlying cognitive processes, our expectations are as follows:

We hypothesize a significant and beneficial effect of annotations compared to no annotations in the VRLE considering the three levels of learning outcome: knowledge, comprehension and application (H1a). The largest positive effect of annotations on learning outcome is expected for knowledge (H1b).

Additionally, the intrinsic motivation in combination with the design of the VRLE might be crucial for learning successfully (Reynolds and Weiner, 2003; Moreno and Mayer, 2007). Therefore, our second research question is (Q2): *Does intrinsic motivation moderate the beneficial effect of annotations on learning outcomes?* As outlined earlier, intrinsic motivation and annotations are both expected to have a beneficial effect on learning outcome. Hence, these two factors might interact in a compensatory way: When learners are motivated, their visual search patterns are efficient and they learn successfully. Adding annotations might build a bridge for a motivational gap and induce a more effective visual pattern for less motivated learners, which should be reflected in an increase of the learning outcome.

We hypothesize that intrinsic motivation significantly moderates the relationship between the treatment factor annotations and learning outcome in the VRLE (H2).

MATERIALS AND METHODS

A Priori Power Analysis

To estimate the necessary sample size, we performed an a priori power analysis. As no previous study investigated signaling in VRLE, we refer to a study that described signaling in a dynamic, system-paced animation for learning biology [De Koning et al., 2010; $f^2(V) = 0.35$]. Based on the meta-analysis of Richter et al. (2016) this study reported rather large effect size (Cohen, 1988). Therefore, we performed our power analysis more conservative [$f^2(V) = 0.31$; $\alpha = 0.05$; power = 0.95]. The needed sample size with the chosen effect size was $N = 60$ (G*Power 3.1.9.4; Faul et al., 2009).

Participants and Design

Our 67 participants were mainly university students in psychology as they were recruited as a part of an undergraduate course in psychology. Due to technical problems, we excluded 6 participants from further analysis. The remaining 61 participants (34.43% male) were aged between 19 and 52 years ($M_{age} = 23.73$; $SD_{age} = 6.00$). About 60% of the participants had at least once been in contact with VR in the past. We applied a between-subject design using a VR app and randomly assigned the participants to one of the two different design options: with ($n = 33$) or without annotations ($n = 28$). As a dependent variable, learning outcome was measured on three different levels: knowledge, comprehension, and application. In addition, learners' prior knowledge and the moderating effect of intrinsic motivation were assessed.

Materials and Measures

In a short *demographic questionnaire*, participants were asked for their gender, age, educational level, their field of studies, and any prior experiences with VR glasses and applications.

Together with a domain-expert, we developed a *pre-test* for prior knowledge that aimed to measure domain-specific knowledge in science, with a special focus on relevant biological and chemical topics for seawater desalination. The test consisted of six open questions (e.g., "What is electrodialysis?") and one closed question (allocation of different particles to particle types), with a maximum score of 20 points. Participants had to complete this knowledge pre-test prior to the learning unit.

To test the effect of annotations in VRLE, the pre-existing German version of the application for smartphones of MOZAIK education (Seawater desalination) was used (mozaWeb3D, 2018). In the respective learning unit, the desalination process for obtaining drinking water from seawater was shown. A special focus was put on different extraction methods and on general information about seawater and its characteristics. The participants saw visualizations of the learning content through a stereoscopic video in VR. Additional information and explanations were offered auditorially. The used application did not allow the learner to interact with and control the

VRLE. The information displayed through the annotations in one of the two experimental conditions did not include further details, but rather simply emphasized certain important aspects of the learning content. The annotations were placed nearby the graphical representation to which they referred to (see **Figure 1**, mozaWeb3D, 2018). The application was run on smartphones that were combined with the VR glass ZEISS VR One Plus.

Furthermore, we developed a *post-test* for measuring the learning outcomes, which consisted of 13 questions on the three levels of Bloom (1956); knowledge, comprehension, application. To measure the learning outcome on knowledge level, six questions were developed including three open and three single choice questions (e.g., “Which process for seawater desalination is particularly environmentally friendly?”). For the comprehension level, four open questions were included (e.g., “Explain why seawater cannot be drunk in large quantities”) and for application we developed three open questions (e.g., “Which everyday object functions similarly to reverse osmosis?”). In the present study, the questions of the post-test aimed to examine different aspects of the learning content. Different questions referred to different concepts and processes. Hence, we expected no high internal consistency of the different questions. To ensure that learning outcome was measured in a rigorous way, we conducted two independent ratings based on a clear evaluation scheme. To analyze the inter-rater reliability, we used the Pearson product moment correlation coefficient, which revealed a very high consistency between the two ratings ($r = 0.92$, $p < 0.001$, $CI = 0.86-0.96$).

Intrinsic Motivation was measured by self-report with the respective subscale of the questionnaire of Lepper et al. (2005), which was based on the original questionnaire of Harter (1981). It consisted of three subscales: challenge, curiosity, and independent mastery. To ensure sufficient quality translation and re-translation has been performed and was compared to the original version. Items were slightly adapted to fit the VR learning context. The internal consistency of the intrinsic motivation scale was sufficient ($\alpha = 0.81$; $CI = 0.74-0.89$).

To gain further insights into the subjective VR experience, the participants were asked to comment their impressions and on their well-being during the VR learning session at the end of the study.

Procedure

The study took place in a realistic learning setting as part of the psychology course at the university. To ensure good data quality without disruption due to technical problems, the study was conducted in a standardized individual testing session with a trained investigator. At the beginning, participants were informed about the procedure of the study and signed an informed consent. All participants started the study by completing the demographic questionnaire and the prior knowledge task *via* an online survey tool (unipark), which took about 15 min. Afterwards, in an individual learning session, learners received a pair of headphones and virtual-reality glasses. The participants watched the VR learning unit about seawater desalination, which lasted about 8 min, and was started by the researcher either with the annotations on or off, to prevent technical problems. After finishing the VR learning unit, the participants filled out a post-test online questionnaire including the test of learning outcome and intrinsic motivation while learning, which took around 15 min. At the end of the questionnaire, participants were able to leave feedback about the learning unit and their subjective experience in VR.

RESULTS

Descriptive Results

Learners' domain-specific prior knowledge was rather low in both experimental conditions (see **Table 1**). The extent of

TABLE 1 | Means and standard deviations of the different experimental conditions.

	With annotations <i>n</i> = 28 <i>M</i> (<i>SD</i>)	Without annotations <i>n</i> = 33 <i>M</i> (<i>SD</i>)
Prior knowledge (%)	23.85 (14.50)	24.45 (14.49)
Intrinsic motivation (<i>max</i> = 15)	8.94 (1.60)	8.76 (1.64)
Learning outcome (%)		
Overall	64.07 (14.87)	57.20 (23.07)
Knowledge	61.29 (19.00)	53.74 (23.43)
Comprehension	64.20 (24.00)	58.80 (28.60)
Application	70.33 (34.33)	62.67 (31.00)

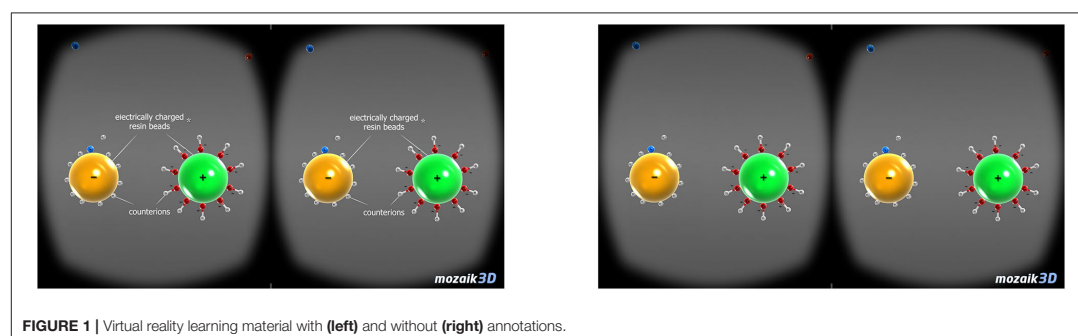


FIGURE 1 | Virtual reality learning material with (left) and without (right) annotations.

intrinsic motivation can be classified as moderate for both groups. The learning outcome was on a medium to a high level, both overall and on the three levels knowledge, comprehension, and application.

No differences between the groups in their preconditions could be found ($F < 1$, $p > 0.203$). Multivariate normal distribution was assumed for the relevant variables for each experimental subgroup ($p > 0.100$). Variances were homogenous based on the Bartlett's-test ($p > 0.060$). For calculation, experimental groups were dummy coded and continuous variables were mean centered. Data preparation and analysis were performed using R 3.5.1.

Effects of Annotations on Learning Outcome

To test our first hypothesis, we analyzed learning outcome based on three of Bloom's (1956) levels. We expected a significant difference between the two experimental groups on all three levels. Analyzing the descriptive pattern of the learning outcome differentiated on the three levels of knowledge, comprehension, and application, we found the expected pattern of our first hypothesis (H1a) on a descriptive level (see **Figure 2**). The learners who learned with annotations outperformed the learners who learned without annotation.

Despite the descriptive patterns the MANOVA did not reveal the expected significant effect of annotations on learning outcome when subcategorized into knowledge, comprehension, and application (H1a; $F < 1$, $p = 0.534$ $\eta^2_{\text{partial}} = 0.037$; see **Table 2**).

Concerning the question of which level of learning processes can be supported best, we compared the effect sizes for the three subcategories, even though none of them were significant (Cohen, 1988). As expected in our hypothesis (H1b), we found the largest effect size for knowledge $\eta^2_{\text{partial}} = 0.031$, which can be classified as a small effect (see **Table 2**). For comprehension and application, we found smaller effect sizes. Even though the descriptive pattern was in line with our expectations, as the experimental groups did not differ significantly in learning outcome (H1a), this hypothesis was not supported by the data.

Effects of Annotations on Learning Outcome Depending on Learners' Intrinsic Motivation

In our second hypothesis, we expected intrinsic motivation to moderate the relationship between the treatment factor annotations and learning outcome. To test this hypothesis, a hierarchical multiple regression analysis was conducted using the mean centered values. In the first step, we included annotations as the treatment factor in the analysis. Additionally, we included intrinsic motivation as the aptitude factor. As theoretically

TABLE 2 | Results of the ANOVA depending on the level of learning outcome.

	$F(1, 59)$	p	η^2_{partial}
Knowledge	1.86	0.177	0.031
Comprehension	0.65	0.425	0.011
Application	0.83	0.366	0.014

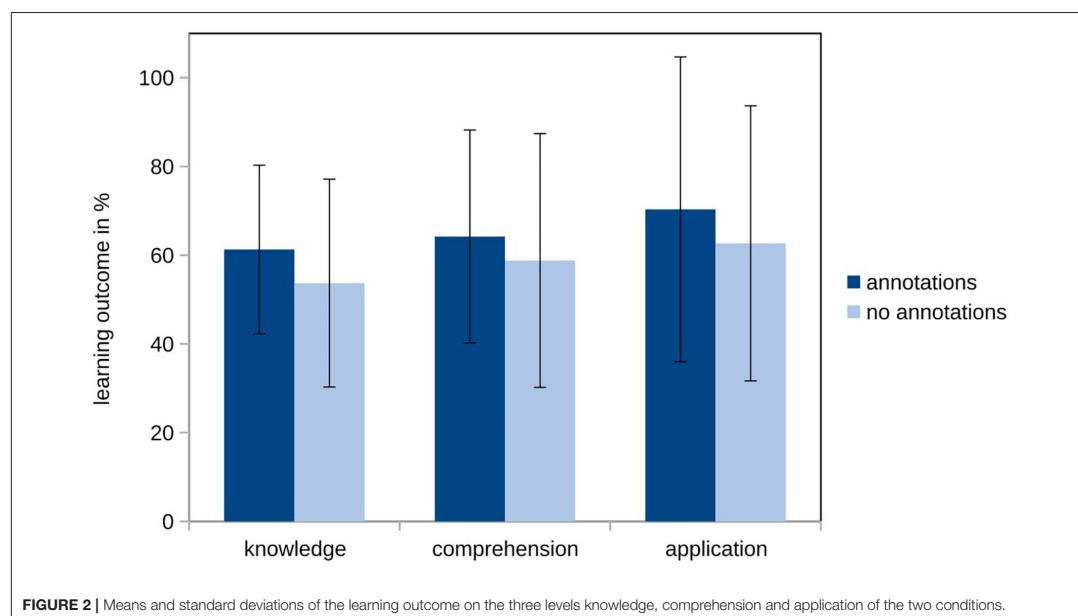


FIGURE 2 | Means and standard deviations of the learning outcome on the three levels knowledge, comprehension and application of the two conditions.

relevant predictor, we included prior knowledge as a control variable but not as an additional moderator, as this was not the focus of the research question. The first model explained $R^2_{adj} = 35.89\%$ of variance in the learning outcome [$F(3, 57)$]

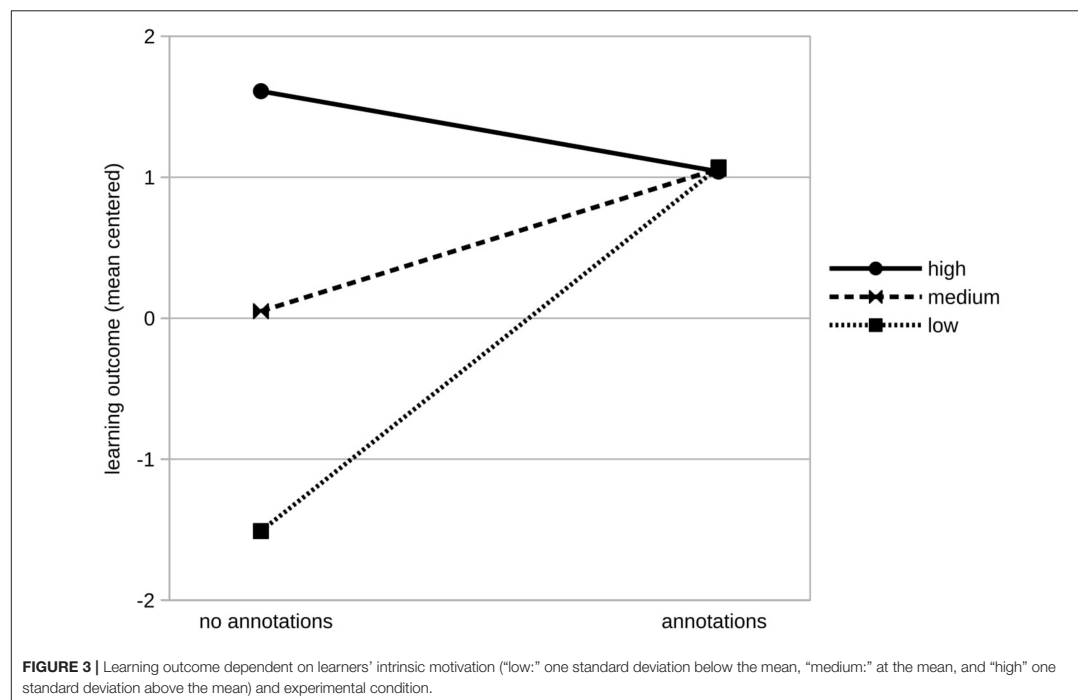
TABLE 3 | Multiple regression model including prior knowledge, intrinsic motivation, and overall learning outcome.

Variable	Estimate	Standard error	t-value	p-value
Model 1				
Intercept	-0.45	0.42	-1.10	0.276
Annotations	1.00	0.62	1.62	0.110
Prior knowledge	0.51	0.11	4.69	<0.001***
Intrinsic motivation	0.53	0.20	2.73	0.008
Model 2				
Intercept	-0.42	0.39	-1.06	0.295
Annotations	1.01	0.59	1.73	0.090
Prior knowledge	0.47	0.11	4.57	<0.001***
Intrinsic motivation	0.97	0.25	3.90	<0.001***
Motivation*Annotations	-0.98	0.37	-2.65	0.011*

*** $p < 0.001$, * $p < 0.05$. Model 1: $R^2_{adj} = 0.36$ $F(3, 57) = 12.20$, $p < 0.001$; Model 2: $R^2_{adj} = 0.42$, $F(4, 56) = 11.86$, $p < 0.001$.

$= 12.20$, $p < 0.001$ ***, see Table 3]. In a second step, we added the aptitude-treatment interaction including annotation and intrinsic motivation to the first model. The second model explained $R^2_{adj} = 41.99\%$ of variance [$F(4, 56) = 11.86$, $f^2 = 0.72$]. Thus, the interaction accounted for a significant proportion of the variance in the learning outcome ($\Delta R^2_{adj} = 6.10$, $p = 0.011$ *, see Table 3).

As expected, prior knowledge had a significant effect on learning outcome ($\beta = 0.47$, $SE = 0.11$, $p < 0.001$). The treatment factor had no significant effect on overall learning outcome ($\beta = 1.01$, $SE = 0.59$, $p < 0.090$) while the aptitude factor intrinsic motivation had a significant effect ($\beta = 0.97$, $SE = 0.25$, $p < 0.001$ ***). We found a significant interaction of annotations and the learners' intrinsic motivation ($\beta = -0.98$, $SE = 0.37$, $p = 0.011$ *). We displayed the moderating effect of intrinsic motivation on learning outcome in Figure 3. To make our result more accessible, we displayed hypothetical groups of learners depending on their intrinsic motivation and experimental group. As recommended by Cohen et al. (2013), we divided learners into three hypothetical groups based on their intrinsic motivation: low ($M - 1$ SD = -1.61), at the mean ($M = 0$), and high ($M + 1$ SD = 1.61; see Figure 3). In the control group in which learners had no support by annotations, learners with high intrinsic motivation outperformed those with lower intrinsic motivation.



For learners in the conditions with annotations, no substantial effect of intrinsic motivation on the learning outcome was found.

Furthermore, we used the Johnson-Neyman technique (Johnson and Fay, 1950) to determine the exact values of learners' intrinsic motivation, for which the conclusion of significant difference between the experimental groups on the learning outcome can be assumed (see Figure 4). For learners with -0.73 or fewer points on the intrinsic motivation score below the mean, the treatment factors annotations had a significant effect on learning success ($p < 0.01$). Hence, learners need either be intrinsically motivated or need annotations to learn successfully.

Virtual Reality Experience While Learning

The participants were asked to comment on the experiment and their subjective impressions in the VRLE. Overall, the participants perceived the VRLE as helpful. The design and explanation were explicitly commended as appealing by 23% of the participants. Additionally, 13% of the participants highlighted that their comprehension of the learning content was fostered by the annotations. Around 30% raised concerns about VR and technical issues because the appropriate equipment is needed to gain high-quality output. Furthermore, some participants (15%) mentioned visual strain while being in VR.

DISCUSSION

While learning complex scientific concepts, learners are usually challenged by integrating multiple representations. This is also the case in a VRLE, where different representations of the

learning content, such as pictorial or textual representations, can be displayed. In the present study, we analyzed whether annotations could support these processes and whether these effects depend on learners' motivation.

Using Annotations to Foster Learning Processes in Virtual Reality

The first hypothesis addressed the effect of annotations on the learning outcome. To gain further insight regarding which cognitive processes are supported best by annotations, we analyzed the effects differentially for the three levels of Bloom (1956): knowledge, comprehension, and application. First, we aimed to answer the question of whether the presence of annotations had a positive effect on the learning outcome on all three levels (H1a). This expectation was not supported by the data. As we predicted in the second part of the first hypothesis, the largest descriptive beneficial effect of annotations was on the knowledge level.

With the merely descriptive, but non-significant, effects, we are in line with previous studies (e.g., Boers et al., 2017). Some previous findings nevertheless described a significant positive effect of annotations on knowledge or recall, but found no effects on higher processing levels, or did not consider them (Mayer and Gallini, 1990; Boers et al., 2017). Other previous findings outline that annotations mainly support deeper cognitive learning processes and learning outcome on the higher levels of comprehension or application (McTigue, 2009; Mason et al., 2013). Taking a closer look at these different results, the question of whether annotations are helpful, and on which level

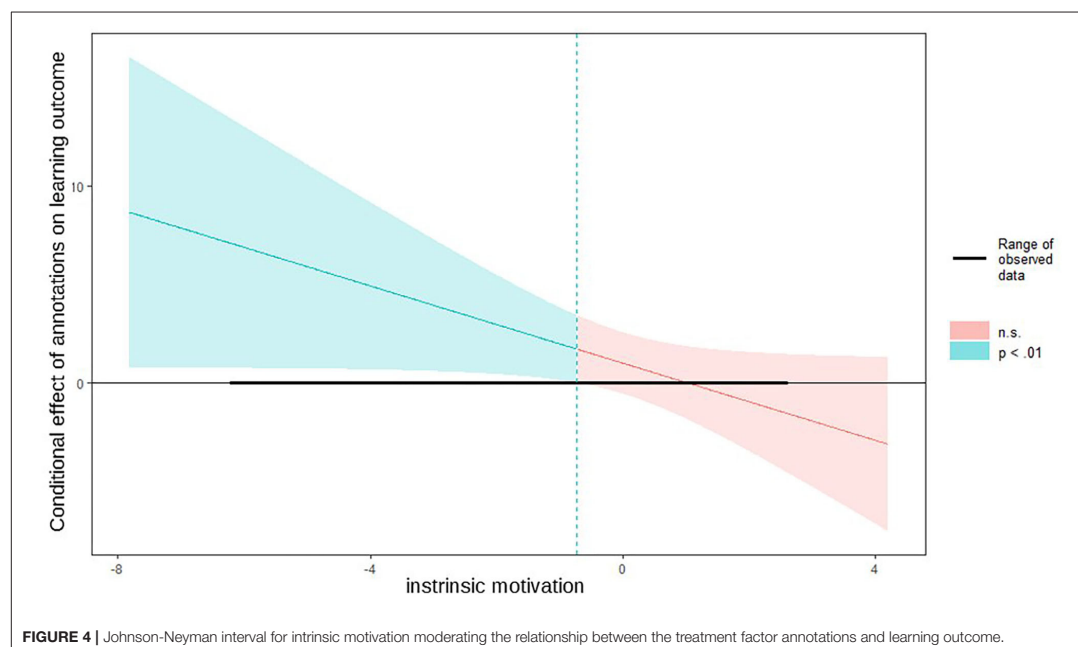


FIGURE 4 | Johnson-Neyman interval for intrinsic motivation moderating the relationship between the treatment factor annotations and learning outcome.

of learning outcome does this beneficial effect occur, depends on the information displayed by the annotation and on the composition of the different representations in the learning material. In the present study, the added annotations were rather simple labels, including one or two words naming relevant components of the learning context. As the labels were placed nearby the corresponding visual information, it also indicated that the corresponding elements could be mapped. However, with this focus on elements, it is plausible that the largest effects have been observed on the knowledge level.

Despite the fact that adding these types of annotations to the VRLE in our present study resulted in a rather small, and not significant, beneficial effect on learning outcome, we carefully outline their impact on the underlying cognitive processes. The positive effects of annotations are assumed to be due to their two functions: signaling and help for element mapping (Gentner et al., 1993; Mayer et al., 1995). The degree to which signaling has a beneficial effect is highly dependent on the visual requirements of the learning material. To uncover a substantial and beneficial effect of annotations, the learning material or environment needs to be sufficiently visually complex (Richter et al., 2016). In the present study, the VRLE consisted of many simplified depictions of the technical or chemical processes. Therefore, the visual display of the learning content was rather a schematic or abstract representation of the learning content. Using abstract representation reduces the unnecessary load while learning, and is hence beneficial (Butcher, 2006). However, this simplification lowers visual requirements, and this might result in lacking the beneficial effects of signaling because they are no longer needed (Richter et al., 2016). As we did not vary the degree of abstraction in our present study, no further assumptions about its influence can be made. Our power analysis was based on a study which also used signaling in a system paced dynamic environment. However, in this study the effect of signaling might be increased as luminance differences were included and by this the effect size of signaling might be enlarged (De Koning et al., 2010).

As outlined earlier, annotations are not *per se* limited to the element layer, but might be extended to foster the relation-to-relation mapping by displaying further information, such as information about related or similar processes of the learning content. Using annotations with relational information might enlarge the positive effect of annotations on the semantic processing and the integration of information into a coherent mental model, and might enlarge the effect on comprehension and application in the learning outcome (Schnotz and Bannert, 2003).

Additionally, in our first hypothesis, we did not consider the learners' aptitudes and states while learning. As learning is an active and resource-consuming process, learners' aptitudes, such as prior knowledge, and the learner's motivational state are crucial for learning success.

Impact of Annotations and Motivation on Learning Outcome

In our second hypothesis, we analyzed the complex interplay of annotations and intrinsic motivation when controlling for prior knowledge. Learners' prior knowledge as a control variable had the expected large impact, and thus our study is in line with many

other studies on expertise-related design effects (Seufert, 2003; Kalyuga, 2009). However, one of the most interesting results of this study is that we found a significant interaction between learners' intrinsic motivation and the presence of annotations. Whereas, learners in the group without annotations differed in their learning outcome depending on their intrinsic motivation, learners in the experimental group with annotations did not show significant differences in their learning outcome depending on their intrinsic motivation. As both factors (intrinsic motivation and signaling through annotations) are described in the literature to have an activating character and a beneficial effect on learning outcomes, our results are in line with these previous findings (Lepper et al., 2005; Richter et al., 2016).

In conclusion, learners are either intrinsically motivated, which results in a higher learning outcome compared to less motivated learners, or they are supported by guiding elements, such as annotations, which reduce visual search while learning. As reflected by the CTML and the CATLM, the learner has an active role while learning (Mayer, 2005; Moreno and Mayer, 2007). Based on the literature, one can assume that when the learner is intrinsically motivated, this goes along with a higher investment of cognitive resources, and this might lead to a higher learning outcome (e.g., Song et al., 2016). As outlined earlier, high intrinsic motivation goes along with more efficient search patterns (e.g., Baranes et al., 2014), and hence learners can compensate for the lack of annotations. Said another way, when less motivated learners received annotations, this also had a beneficial effect on learning, presumably by guiding attention and reducing visual search.

The findings of the present study imply that using annotations in a VRLE has a beneficial effect on the learning outcome, and might bridge the motivation gap by fostering learners who are not intrinsically motivated. But, of course, one could ask why the annotations did not work as an enhancer for motivated students, which would have resulted in even higher scores (i.e., a synergetic effect). In our study, the visual requirements of the VRLE were assumed to be rather low. A twofold activation, leading to "double" investment, seems to have not been necessary. Therefore, a synergetic effect of intrinsic motivation and annotations was neither expected nor found. However, this might be questionable because learners did not reach the maximum of learning outcome in neither of the two activating conditions. Like an illusion of knowing (Avhustiuk et al., 2018), learners might have had the (false) impression that they reached their maximum, and invested sufficient resources. So, one might ask what would have been necessary to activate learners incrementally. Nevertheless, the importance of considering the learners' aptitudes and their current state while learning when analyzing the effects of VRLE was underlined by the present findings.

Strength and Weaknesses and Recommendations for Future Studies

As the results of the present study indicate, an activating effect of annotations or increased intrinsic motivation on the learner, the underlying mechanisms and cognitive processes need to be further explored. In our present study, the main focus was to uncover the effect of adding annotation and interaction with

intrinsic motivation while learning. Moreover, further cognitive mechanisms might be crucial and need to be taken into account. For instance, to uncover the perceived cognitive demands, which are posed by the VRLE, future studies should include a differentiated measurement of cognitive load. Such a measure could help to gain further insight into the distribution of the learners' cognitive resources with regard to intrinsic, extraneous, and germane loading aspects (Klepsch et al., 2017). On the one hand, this would lead to further insights into whether or not additional annotations increase the cognitive demands, particularly the extraneous cognitive load. On the other hand, annotations or motivation might be activating and increase the invested resources of the learner while learning, which would be reflected by germane cognitive load. Additionally, collecting process data *via* eye-tracking or pupillometry might enable deeper insight into the attention guiding effects of annotations and their effects on visual search, as well as the visual requirements of the learning material. Furthermore, the visual search pattern of the integration of information of multiple representations while learning might be investigated by this additional data.

In the present study, we focused on measuring intrinsic motivation. Further aspects of motivation might also be explored in future studies. In our study, motivation has only been treated as a state variable. Motivational traits, for example, goal orientation, might also have a substantial impact on the learning outcome and need to be researched in the field of VRLE (Wolters et al., 1996; Eccles and Wigfield, 2002).

Despite the fact that we found a positive and compensating effect of annotations in VRLEs, the effect might be highly dependent on the concrete design of the learning material, learning content as well as the chosen VRLE. VRLEs range from non-immersive 3D computer simulations to immersive VR with 6 degrees of freedom. Especially in an applied setting such as university courses, the use of simple VR headset like we used in our study would be easier to implement compared to high fidelity VR headsets. However, future studies might investigate the beneficial effects of annotations with other VRLEs. One limitation of this study is that we used a VRLE that neither allows the learners to choose their own perspective nor allows them to interact with the virtual environment. Hence, the field of view was determined. This goes along with the beneficial effect that the challenge of guiding attention in the VRLE does not have to be faced. Therefore, only attentional processes in the field of view need to be considered. When learners are able to choose their own perspective in VR, this leads to additional challenges for the design of VRLEs, and makes it necessary to monitor the learners by using process data during their VR learning session.

The positive effect of annotations for learners with low motivation can therefore not be generalized for all kinds of VRLEs, as the interactivity in the VR might have an impact on learning outcome. Interaction in VRLEs might also have a beneficial effect on motivational processes, but might cause additional task load or cognitive load, and therefore might be distracting. Hence, further studies are required to gain insights into the potential of VRLEs under consideration of cognitive processes and motivational aspects. Additionally, the setting of the VRLE needs to be considered. In our study, learners

used the VRLE in a controlled setting. As uncovered in the pilot study, learners are rather challenged by the handling of the unfamiliar VR-application, and therefore the researcher had to assist with the technical setup. In line with this idea, participants raised many concerns about technical problems in their subjective feedback. Therefore, the question can be raised whether using this kind of learning environment is possible in real-life settings, such as university courses or classroom where the implementation of individual support would be challenging (Checa and Bustillo, 2020). One approach to this challenge would be to train the learners to interact successfully with the technical equipment, and to foster their handling skills in general. This would enhance their ease of use and would foster their technical skills, which could be necessary for future work settings.

CONCLUSION

New possibilities to teach scientific concepts in VRLEs come along with the challenge to choose an appropriate design. An adequately designed VRLE supports learners in integrating the learning content into a coherent mental model. For concrete visualization and implementation, learners' aptitudes, for instance, prior knowledge and intrinsic motivation while learning, need to be considered. A lack of intrinsic motivation may be bridged by adding supportive elements, such as annotations. Whereas, highly intrinsically motivated learners seem to successfully face the challenge of integrating information in the mental model, learners who lack intrinsic motivation need to be supported. Future research should focus on uncovering further details about motivational and cognitive processes while learning in VRLEs. Particularly, motivational and cognitive processes, in combination with elements guiding attention and supporting coherence formation, may be essential. One future challenge is to uncover the possibilities of VRLEs and to deduce certain design recommendations and principles for VRLEs.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AV and PA contributed to the conception, design of the studies, developed the used questionnaires, and led data collection for all studies. TS revised the questionnaires. AV analyzed and interpreted the data and drafted the work, which was revised critically by PA and TS. All authors provided approval of the final submitted version of the

manuscript, agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy, or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

- Adesope, O. O., and Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: a meta-analysis. *J. Educ. Psychol.* 104:250. doi: 10.1037/a0026147
- Akçayır, M., Akçayır, G., Pektaş, H. M., and Ocak, M. A. (2016). Augmented reality in science laboratories: the effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. *Comput. Hum. Behav.* 57, 334–342. doi: 10.1016/j.chb.2015.12.054
- Avhustiuk, M. M., Pasichnyk, I. D., and Kalamazh, R. V. (2018). The illusion of knowing in metacognitive monitoring: effects of the type of information and of personal, cognitive, metacognitive, and individual psychological characteristics. *Eur. J. Psychol.* 14, 317–341. doi: 10.5964/ejop.v14i2.1418
- Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359
- Baranes, A. F., Oudeyer, P.-Y., and Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Front. Neurosci.* 8:317. doi: 10.3389/fnins.2014.00317
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York, NY: David McKay Co Inc.
- Boers, F., Warren, P., Grimshaw, G., and Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Comput. Assist. Lang. Learn.* 30, 709–725. doi: 10.1080/09588221.2017.1356335
- Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *Internet Higher Educ.* 33, 24–32. doi: 10.1016/j.iheduc.2017.01.004
- Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. *J. Educ. Psychol.* 98, 182–197. doi: 10.1037/0022-0663.98.1.182
- Checa, D., and Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools Appl.* 79, 5501–5527. doi: 10.1007/s11042-019-08348-9
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Routledge. doi: 10.4324/9780203774441
- De Koning, B. B., Tabbers, H. K., Rikers, R. M., and Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Appl. Cogn. Psychol.* 21, 731–746. doi: 10.1002/acp.1346
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., and Paas, F. (2010). Learning by generating vs. receiving instructional explanations: two approaches to enhance attention cueing in animations. *Comput. Educ.* 55, 681–691. doi: 10.1016/j.compedu.2010.02.027
- Di Natale, A. F., Repetto, C., Riva, G., and Villani, D. (2020). Immersive virtual reality in K-12 and higher education: a 10-year systematic review of empirical research. *Br. J. Educ. Technol.* 51, 2006–2033. doi: 10.1111/bjet.13030
- Di Serio, Á., Ibáñez, M. B., and Kloos, C. D. (2013). Impact of an augmented reality system on students' motivation for a visual art course. *Comput. Educ.* 68, 586–596. doi: 10.1016/j.compedu.2012.03.002
- Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Ann. Rev. Psychol.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153
- Eseryel, D., Law, V., Ifenthaler, D., Ge, X., and Miller, R. (2014). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Educ. Technol. Soc.* 17, 42–53.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Fowler, C. (2015). Virtual reality and learning: where is the pedagogy? *Br. J. Educ. Technol.* 46, 412–422. doi: 10.1111/bjet.12135
- Fransson, A. (1977). On qualitative differences in learning: effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *Br. J. Educ. Psychol.* 47, 244–257. doi: 10.1111/j.2044-8279.1977.tb02353.x
- Gentner, D., Rattermann, M. J., and Forbus, K. D. (1993). The roles of similarity in transfer: separating retrievability from inferential soundness. *Cogn. Psychol.* 25, 524–575. doi: 10.1006/cogp.1993.1013
- Han, I. (2020). Immersive virtual field trips in education: a mixed-methods study on elementary students' presence and perceived learning. *Br. J. Educ. Technol.* 51, 420–435. doi: 10.1111/bjet.12842
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: motivational and informational components. *Dev. Psychol.* 17, 300–312. doi: 10.1037/0012-1649.17.3.300
- Johnson, P. O., and Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika* 15, 349–367. doi: 10.1007/BF02288864
- Kalyuga, S. (2009). *Managing cognitive load in adaptive multimedia learning*. Hershey, Pa: IGI Global. doi: 10.4018/978-1-60566-048-6
- Kalyuga, S., and Sweller, J. (2014). "The redundancy principle in multimedia learning," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*, ed R. E. Mayer (Cambridge, New York, NY: Cambridge University Press), 247–262.
- Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.1017/CBO9781139547369.013
- Klingenberg, S., Jørgensen, M. L., Dandanell, G., Skriver, K., Mottelson, A., and Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *Br. J. Educ. Technol.* 51, 2115–2138. doi: 10.1111/bjet.13029
- Lepper, M. R., Corpus, J. H., and Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: age differences and academic correlates. *J. Educ. Psychol.* 97, 184–196. doi: 10.1037/0022-0663.97.2.184
- Makransky, G., Borre-Gude, S., and Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *J. Comput. Assist. Learn.* 35, 691–707. doi: 10.1111/jcal.12375
- Mason, L., Pluchino, P., and Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: evidence from eye movements. *Reading Res. Quart.* 48, 199–214. doi: 10.1002/rrq.41
- Mautone, P. D., and Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *J. Educ. Psychol.* 93, 377–389. doi: 10.1037/0022-0663.93.2.377
- Mayer, R. E. (2005). "Principles for reducing extraneous process in multimedia learning: Coherence, signals, redundancy, spatial contiguity, and temporal contiguity principles," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*, ed R. E. Mayer (Cambridge, New York, NY: Cambridge University Press), 325–337. doi: 10.1017/CBO9780511816819.013
- Mayer, R. E. (2014). *The Cambridge Handbook of Multimedia Learning*. Cambridge, New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139547369
- Mayer, R. E., and Fiorella, L. (2014). "12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles," in *Cambridge Handbooks in Psychology. The Cambridge Handbook of Multimedia Learning*, ed R. E. Mayer (Cambridge, New York, NY: Cambridge University Press), 279–309. doi: 10.1017/CBO9781139547369.015

- Mayer, R. E., and Gallini, J. K. (1990). When is an illustration worth ten thousand words? *J. Educ. Psychol.* 82, 715–726. doi: 10.1037/0022-0663.82.4.715
- Mayer, R. E., Steinhoff, K., Bower, G., and Mars, R. (1995). A generative theory of textbook design: using annotated illustrations to foster meaningful learning of science text. *Educ. Technol. Res. Dev.* 43, 31–41. doi: 10.1007/BF02300480
- McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemp. Educ. Psychol.* 34, 143–153. doi: 10.1016/j.cedpsych.2008.12.003
- Mikropoulos, T. A., and Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Comput. Educ.* 56, 769–780. doi: 10.1016/j.compedu.2010.10.020
- Miranda, A. T., and Palmer, E. M. (2014). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behav. Res. Methods* 46, 159–172. doi: 10.3758/s13428-013-0357-7
- Moreno, R., and Mayer, R. (2007). Interactive multimodal learning environments. *Educ. Psychol. Rev.* 19, 309–326. doi: 10.1007/s10648-007-9047-2
- mozaWeb3D (2018). Seawater Desalination [video Game]. Szeged: Mozaik Education Ltd. Available online at: https://www.mozaweb.com/de/Extra-3D_Modelle-Die_Meerwasserentsalzung-210610 (accessed October 31, 2019).
- Ozcelik, E., Arslan-Ari, I., and Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Comput. Hum. Behav.* 26, 110–117. doi: 10.1016/j.chb.2009.09.001
- Paes, D., Arantes, E., and Irizarry, J. (2017). Immersive environment for improving the understanding of architectural 3D models: comparing user spatial perception between immersive and traditional virtual reality systems. *Automat. Constr.* 84, 292–303. doi: 10.1016/j.autcon.2017.09.016
- Paivio, A. (1990). *Mental Representations: A Dual Coding Approach*. Cary: Oxford University Press, Incorporated. doi: 10.1093/acprof:oso/9780195066661.001.0001
- Parong, J., and Mayer, R. E. (2018). Learning science in immersive virtual reality. *J. Educ. Psychol.* 110, 785–797. doi: 10.1037/edu0000241
- Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., et al. (2016). Virtual laboratories for education in science, technology, and engineering: a review. *Comput. Educ.* 95, 309–327. doi: 10.1016/j.compedu.2016.02.002
- Radianti, J., Majchrzak, T. A., Fromm, J., and Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. *Comput. Educ.* 147:103778. doi: 10.1016/j.compedu.2019.103778
- Rapp, D. N. (2007). "Mental models: theoretical issues for visualizations in science education," in *Models and Modeling in Science Education: v. 1. Visualization in Science Education*, Vol. 255, ed G. J. K. (Dordrecht: Springer), 43–60. doi: 10.1007/1-4020-3613-2_4
- Renkl, A., and Scheiter, K. (2017). Studying visual displays: how to instructionally support learning. *Educ. Psychol. Rev.* 29, 599–621. doi: 10.1007/s10648-015-9340-4
- Reynolds, W. M., and Weiner, I. B. (2003). *Educational Psychology. Handbook of Psychology*, Vol. 7. Hoboken, NJ: Wiley.
- Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–36. doi: 10.1016/j.edurev.2015.12.003
- Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020
- Schnotz, W., and Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learn. Instr.* 13, 141–156. doi: 10.1016/S0959-4752(02)00017-8
- Schott, C., and Marshall, S. (2018). Virtual reality and situated experiential education: a conceptualization and exploratory trial. *J. Comput. Assist. Learn.* 34, 843–852. doi: 10.1111/jcal.12293
- Schüler, A., Arndt, J., and Scheiter, K. (2015). Processing multimedia material: does integration of text and pictures result in a single or two interconnected mental representations? *Learn. Instr.* 35, 62–72. doi: 10.1016/j.learninstruc.2014.09.005
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learn. Instr.* 13, 227–237. doi: 10.1016/S0959-4752(02)00022-1
- Seufert, T., Jänen, I., and Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Comput. Hum. Behav.* 23, 1055–1071. doi: 10.1016/j.chb.2006.10.002
- Song, H. S., Kalet, A. L., and Plass, J. L. (2016). Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. *J. Comput. Assist. Learn.* 32, 31–50. doi: 10.1111/jcal.12117
- Wilkerson-Jerde, M. H., Gravel, B. E., and Macrander, C. A. (2015). Exploring shifts in middle school learners' modeling activity while generating drawings, animations, and computational simulations of molecular diffusion. *J. Sci. Educ. Technol.* 24, 396–415. doi: 10.1007/s10956-014-9497-5
- Wolters, C. A., Yu, S. L., and Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learn. Individ. Differ.* 8, 211–238. doi: 10.1016/S1041-6080(96)90015-1
- Wu, B., Yu, X., and Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: a meta-analysis. *Br. J. Educ. Technol.* 51, 1991–2005. doi: 10.1111/bjet.13023
- Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *J. Comput. Assist. Learn.* 23, 120–132. doi: 10.1111/j.1365-2729.2006.00215.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Vogt, Albus and Seufert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The second paper was published in the Journal Computers & Education under the terms of the Creative Commons CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model. *Computers & Education*, 171, 104235.

<https://doi.org/10.1016/j.compedu.2021.104235>

Journal Metrics: 2-Year Impact Factor: 8.538 (2020), h-index 179



Contents lists available at ScienceDirect

Computers & Education

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

Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model

Andrea Vogt^{a,*}, Franziska Babel^b, Philipp Hock^b, Martin Baumann^b, Tina Seufert^a

^a Ulm University, Institute of Psychology and Education, Department Learning and Instruction, Albert-Einstein-Allee 47, 89081, Ulm, Germany

^b Ulm University, Institute of Psychology and Education, Department Human Factors, Albert-Einstein-Allee 43, 89081, Ulm, Germany

ARTICLE INFO

Keywords:

Virtual reality
Multimedia learning
Elaboration prompt
Mental models
Structural assessment of knowledge

ABSTRACT

Immersive virtual reality learning environments (VRLEs) have great potential but they do not guarantee effective or in-depth learning. Activating learners by including prompts, has been reported to increase learning in VR. Based on their promising effects in conventional settings, elaboration prompts should increase learning outcomes by activating learners to explain the learning content in their own words or to contrast it to prior knowledge. In our study, we investigated the effect of an elaboration prompt before entering the VRLE on learning robotics. Our between-subject study (prompt vs. no prompt) included $N = 64$ participants. We applied a multi-method approach (functional and structural) to measure mental models. We used a differentiated learning outcome measurement (knowledge, comprehension, and application) as a functional measurement. By using the structural assessment of knowledge technique, we deduce the structural accuracy of the developed mental model. We found an increase by the elaboration prompt on knowledge level but not on comprehension or application level. The interplay of pre-accuracy and prompt had a significant impact on structural accuracy after learning. Application level and structural accuracy after learning were positively correlated. Our multi-method approach delivered insights on different aspects of the developed mental model when prompting before entering the VRLE. For future research, further approaches to explore in-depth learning in VRLEs are recommended.

1. Introduction

Learning scientific content or technical concepts is often abstract, not intuitive or easy to access. Using virtual reality learning environments (VRLEs) is one promising approach to depict technical concepts for higher education learning (Wu, Yu, & Gu, 2020). Virtual reality (VR) is defined as 'a computer-mediated simulation that is three-dimensional, multisensory, and interactive, so that the user's experience is 'as if' inhabiting and acting within an external environment' (Burbules, 2006, p. 37). For instance, VRLEs have been used to create situated and realistic learning environments where learners can experience micro or macro phenomena otherwise hard or impossible to include in an everyday university setting (Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020; Calvert & Abadia, 2020). When learning with simulations in VR, this allows learners to experience the learning content in a multisensory way. Particularly, by using immersive VR, including a head-mounted display with a stereoscopic view, the VRLE is experienced from the inside and offers the possibility to fully immerse oneself into the learning environment (Cummings & Bailenson, 2016).

* Corresponding author.

E-mail address: andrea.vogt@uni-ulm.de (A. Vogt).

<https://doi.org/10.1016/j.compedu.2021.104235>

Received 12 January 2021; Received in revised form 14 April 2021; Accepted 10 May 2021

Available online 16 May 2021

0360-1315/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Many previous studies aimed to explore the potential of VRLEs and thus compared learning in immersive VR with desktop VR or conventional learning material (Howard & Lee, 2020; Klingenberg et al., 2020; Parong & Mayer, 2020). Heterogeneous effects were reported: The meta-analysis of Wu et al. (2020) summarized 35 studies investigating the effects of immersive VR. The overall effect indicated that learning in immersive VR was more effective than other learning settings. However, they reported that 34% of the studies showed no or even adverse effects. For instance, Makransky, Andreassen, Baceviciute, and Mayer (2020) described the beneficial effects of learning in immersive VR on liking but not on learning outcome. Moreover, Parong and Mayer (2020) describe that learning with powerpoint slides was more effective than learning in immersive VR. In addition, Baceviciute, Mottelson, Terkildsen, & Makransky (2020), describe that reading in VR led to a higher retention than listening to an auditory text. Based on prior findings, one can conclude that there are many influencing factors for learning successfully in VR: the learning content and its presentation, the instructional design, and the learners' prior knowledge and abilities.

Hence, VRLEs do not per se guarantee successful learning. Therefore, the question arises how to support learners to turn the potential of VRLEs into an actual in-depth learning approach. Generally, there are two possible ways for such an approach: on the one hand, learners might be supported by providing relevant information as pre-training before learning, i.e. the support specifies *what to learn* (Meyer, Omdahl, & Makransky, 2019). On the other hand, learners might be supported on a strategic level, i.e. the support specifies *how to learn*. Parong and Mayer (2018) investigated one approach of strategic support: They asked learners to summarize the learning content after each VR learning session, which resulted in a beneficial effect on learning outcome. However, summary prompts means that learners must have an additional workstation and may have to switch back and forth between immersive VR (which takes a larger space to interact) and their desk. Depending on the implementation and the given learning situation, this can be an organizational challenge and rather costly.

Therefore, we aimed to investigate an approach that includes the advantages of both approaches by including an elaboration prompt, before entering the immersive VRLE. Thus, we offer strategic support, which can be easily generalized to different learning contexts and environments, but which is nevertheless provided beforehand not to interrupt the learning process in the VRLE. *But, what is an elaboration prompt?* In general, prompts aim to activate and stimulate learners either on a meta-cognitive or cognitive level to foster related learning processes (Bannert & Reimann, 2012; Gentner & Seufert, 2020). More precisely, *cognitive prompts* can be used to activate cognitive learning processes in a targeted manner. This can then result in a more in-depth elaboration of the learning content. As this is what we are aiming to, we use cognitive prompts that particularly foster in-depth elaboration of the learning content. These prompts are also referred to as *elaboration prompts* (e.g. Endres, Carpenter, Matrin, & Renkl, 2017; Loibl & Leuders, 2019; further details see 3.3).

We used a multi-method approach to explore the positive effects of our elaboration prompt on learning. Particularly, we investigated the prompt's effect on the development of a mental model containing abstract technological concepts. On the one hand, our focus was on learners' performance or more precisely on the extent to which learners are able to answer questions on different levels of learning outcome (knowledge, comprehension, application; Bloom, 1956). This can be described as a functional approach to measure mental models. Furthermore, we measured which knowledge structure was given before and after the learning unit. By using the *structural assessment of knowledge* technique, we determined learners' structural accuracy compared with the respective counterpart, based on expert ratings (Trumpower & Vanapalli, 2016; for further details see appendix A). Furthermore, as the development of a mental model also depends on learners' aptitudes, we also considered these in our study.

2. Mental models and learning in virtual reality

To display the learning content in VR, different information sources are often combined (e.g. Parong & Mayer, 2018). The beneficial effect of combining pictorial and textual components (multimedia effect) has been well researched and established when learning with conventional material (Mayer, 2009). Therefore, using multiple and different representations might also be promising in VRLEs. When learning successfully, this information is built up to an accurate, coherent mental model (Mayer, 2014; Scheiter, Schüler, & Eitel, 2017). On the one hand, VRLEs offer substantial learning opportunities because learners are able to choose their preferred position to process the textual and pictorial representations. Furthermore, learners are able to explore the learning content by interacting with the given objects in the virtual world. On the other hand, this might result in simply watching the learning material superficially. Hence, to understand potential challenges while learning in VR, the underlying cognitive processes need to be considered.

The *Cognitive Theory of Multimedia Learning* (CTML) by Mayer (2014) offers insight into the perception, processing, and integration of learning content. This theory describes processing depending on different codes (text or picture) or different modalities (visual or auditory; Paivio, 1990). If learning content is presented in different codes, this goes along with processing via both channels and the probability of successful memory retrieval is increased (Paivio, 1990). For instance, while learning in VR, spoken text, animations, or pictures of the learning content are combined to trigger this beneficial effect (e.g., Parong & Mayer, 2018; Radianti et al., 2020). While processing the learning content, first, two separate models, the verbal and the pictorial model, are built from the different information sources (Mayer, 2014). At the end of the learning process, these two mental models are integrated into one coherent mental model. While processing both verbal and pictorial information building one coherent mental model is the main goal. Johnson-Laird (1981) described *mental models* as internal models of the world which are analog in nature and are set up based on perception and information processing. Based on these developed mental models, learners are able to answer questions on different skill levels.

To understand the integration process of texts and pictures into the mental model, the *Integrated Model of Text and Picture Comprehension* (IMTPC) provides more details (Schnotz & Bannert, 2003). Because pictures are analog in nature, no additional translation process is required after processing the information semantically to include them into their analog counterpart, the mental

model (Schnotz, Böckheler, & Grzondziel, 1999). In line with other text comprehension models (e.g. Newmeyer, 1996) the IMTPC describes an additional translation process when including verbal information (written or spoken) into the mental model. After processing sub-semantically, an internal representation results based on the given text syntax. When verbal information is then semantically processed, this results in propositional representations (Schnotz & Bannert, 2003). To include these propositional representations into the analog mental model, these need to be translated. Hence, there is a fundamental distinction between propositional representations and mental models. Propositional representations are described as strings of symbols with an arbitrary syntactic structure while mental models are analog in nature (Johnson-Laird, 1981). Furthermore, the IMTPC describes a possible shortcut between the internal text surface representation and the mental model. Hence, it assumes this superficially processed textual information can also be stored and integrated into the mental model without being further processed semantically (Schnotz & Bannert, 2003).

2.1. Measuring mental models

To answer the question of how to measure a mental model, these need to be first characterized in more detail. As outlined earlier, mental models represent an external world's internal model (Johnson-Laird, 1981). Although there is a basic consensus about mental models' analog nature, different definitions emphasize the mental model's more specific characteristics (Hong & O'Neil, 1992). For instance, Rouse and Morris (1986) described mental models as a specific type of knowledge in the respective field of interest in contrast to a more general view of knowledge. White and Frederiksen (1990) defined mental models as knowledge structures that incorporate different types of knowledge: declarative, procedural, and a control structure. Another definition described the term mental model to be connected with the organization and structure of the processed knowledge (Kraiger, Ford, & Salas, 1993). Hence, based on these mental models, certain functions of concepts can be deduced and future events anticipated (Rouse & Morris, 1986). Consequently, when learning in VR different aspects of the concept mental model can be used as indicators for learning success. The mental model's resulting structure represents a more abstract part of the mental model, giving information about relationships and distances between the included concepts without relying on direction or describing concrete consequences. When measuring mental models by a multi-method approach, this might uncover different aspects of mental models: On the one hand different levels of information processing of declarative knowledge (semantic or sub-semantic) might be measured by a *functional* approach. On the other hand, the resulting mental model's configuration or organization structure including distances and relationships between different concepts might be uncovered using a *structural* approach (Kraiger et al., 1993).

Therefore, both functional and structural approaches were chosen to assess the quality of learners' mental models in our study. To measure mental models' functional aspects, performance tests with questions differentiating sub-semantic and semantic processing are used. Bloom's (1956) first level (knowledge) reflects sub-semantic processing and thus learners' performance on simple recall of terms or definitions. In contrast, higher levels (comprehension and application) require semantic processing (Bloom, 1956). To answer comprehension questions, learners must predict the effect of manipulating one component onto another. For application, a more global understanding is necessary. Different components, their relating cause-and-effect chains, and their connection to already stored knowledge must be correctly combined and applied to answer these questions. Using this taxonomy leads to insights into the connection of the concepts on a functional level but does not deliver insights into the structure of the mental model itself.

To analyze the mental models' structure, the *structural assessment of knowledge* technique (SAK; Kraiger et al., 1993) can be used. SAK is a summative assessment to investigate the resulting knowledge network's accuracy and its undirected connections and distance between included concepts (Jonassen, 1995). The SAK and the resulting structural accuracy are specified based on the rating of pairs of concepts for their undirected connectedness (Trumpower & Vanapalli, 2016). These ratings can be compared to experts' knowledge networks and accuracy can be deduced (Acton, Johnson, & Goldsmith, 1994). For instance, the expert might rate the relationship between the concepts *actuator technology* and *sensor technology* as with a 2, standing for a rather weak relationship and a high distance between these concepts. When participants rate these concepts to be strongly related in contrast to those of the experts, this is reflected in a low accuracy score (for further details see appendix A). This technique has been successfully used in the field of learning science to examine the structure of the resulting mental model (Trumpower & Vanapalli, 2016; Wouters, van der Spek, & van Oostendorp, 2011).

2.2. Fostering in-depth learning processes by using prompts

The theory of *generative learning activities* described by Fiorella and Mayer (2016) refers to an active learning process, including making sense of the learning content by integrating new information into the pre-existing mental model. Prompts can support generative learning activities, for instance, when they ask learners to learn by summarizing (e.g. Parong & Mayer, 2018). *Prompts* are described as strategy activators because they support specific cognitive strategies (Reigeluth, 1983; Bannert & Reimann, 2012). Hence, they increase learning outcomes and support the development of a mental model (Bannert & Reimann, 2012; Berthold, Eysink, & Renkl, 2009; Berthold, Nückles, & Renkl, 2007). Additionally, prompts might enable learners to transfer what they have learned to new contexts.

Studies have described different realizations of prompts in the learning context. For instance, learners were asked to take notes, think of examples of their everyday life, or conduct short summaries (e.g., Bannert & Reimann, 2012; Parong & Mayer, 2018; Endres et al., 2017). All these prompts aim to foster in-depth in contrast to superficial and non-semantic learning processes and they activate the elaboration of the learning content. Therefore, these are also named *elaboration prompts*. In one study, learners were activated to connect and compare relevant concepts of the learning content. Learners in the elaboration prompt group outperformed the learners in the control group (Endres et al., 2017). Including explanatory components, which activate the learners to process the content deeply

and rephrase the relevant concepts in their own words, is particularly effective (Berthold et al., 2009; for further details see 3.3).

Few studies have examined the effects of prompting in VRLEs. In this context, learners were encouraged to interact with a VRLE in which learning content was repeated and important concepts were outlined by the prompts (Kamarainen et al., 2018), or learners were asked to interrupt the VR experience to write short summaries (Parong & Mayer, 2018). The beneficial effects of these prompts on learning outcome were described and uncovered in simple recall and higher learning outcomes. These prompts have in common that they were included in the learning unit and require specific adaption to the respective learning unit. Reflecting on these findings, one promising and very flexible approach to activate learners is to use *elaboration prompts*, including an explanatory component before entering the VRLE. As outlined before, when adding the elaboration prompt into the instruction, these do not have to be adapted to the specific VR learning content and do not interrupt the learning process and are very flexible to implement.

Our study investigated this beneficial effect of using an elaboration prompt before entering the immersive VRLE. To provide insights into the resulting mental model, we used the aforementioned functional and structural approaches: learning outcome measures on different levels and structural accuracy measured by SAK. Furthermore, we considered specific aptitudes of the learners to control their impact on developing a mental model. Pre-existing prior knowledge influenced the mental model's development and learning outcome (e.g., learning outcome based on Blooms' taxonomy: Seufert, 2019; Lehmann & Seufert, 2018, and structural: Riemer & Schrader, 2016; Trumpower & Vanapalli, 2016). Therefore, prior knowledge might impact the beneficial effect of the prompt. Furthermore, learners' abilities, namely verbal ability, was described as crucial for the verbalization of the developed mental model: The influence of verbal ability on different learning outcome levels (knowledge, comprehension, and application) is well proven (Wallen, Plass, & Brünken, 2005). As verbal ability has been described to impact accuracy scores significantly, we also considered this aptitude in relation to our structural approach to measuring mental models (Jonassen, 1995).

2.3. Present study

While learning in VR, learners are challenged by processing and integrating all relevant information in one coherent mental model. Therefore, we implemented an elaboration prompt to foster the semantic processing of the learning content (Bannert & Reimann, 2012; Endres et al., 2017). Hence, we expected the elaboration prompt to particularly impact the learning outcome's higher levels because these reflect semantic processing. Our expectations were as follows:

- We hypothesized that a significantly higher learning outcome would be observed for the group with prompt than in the group without prompt for all three levels of Bloom (1956; knowledge, comprehension, application; H1a). Additionally, we expected the elaboration prompt to cause a larger beneficial effect on the higher learning outcomes, namely, comprehension and application, compared to the knowledge level (H1b).

Furthermore, we investigated whether the elaboration prompt had beneficial effects on the structural accuracy of the resulting mental model (Kraiger et al., 1993; Trumpower & Vanapalli, 2016). As building connections is highly dependent of the level and sophistication of pre-existing knowledge networks, we assume that the beneficial effect of prompting on the post-accuracy score depends on the pre-accuracy score (Kraiger et al., 1993; Scheiter et al., 2017; Trumpower & Vanapalli, 2016). Hence, we expect the following effect:

- We hypothesized that the elaboration prompt's beneficial effect on structural post-accuracy of learners' mental model depends on the pre-accuracy (H2).

Previous findings outlined, that for both, high structural accuracy and answering questions on the application level require a more global understanding of the connectedness of the different components, compared to the knowledge and comprehension level, these two measurements are expected to be related (Bloom, 1956; Davis, Curtis, & Tschetter, 2003; Kraiger et al., 1993; Trumpower & Vanapalli, 2016). For instance, Goldsmith, Johnson, and Acton (1991) described a positive relationship between structural accuracy and learning outcome. Thus, based on theoretical assumptions and prior findings we assume as follows:

- We hypothesized a significant positive correlation between the mental model's accuracy and learning outcome based on the application level (H3).

To gain further insights into the processes while learning in VR, we investigate as an additional exploratory research question, how the elaboration prompt affects learners' distribution of attention in the VRLE. To quantify the learners' attention, we explored the time the robot was focussed in the learners' field of view. As our analysis was based on a very brief measurement of attention based on head-tracking data, no specific hypothesis was postulated.

3. Material and method

3.1. A priori power analysis

To determine the necessary sample size, an a priori power analysis was performed. To estimate the effect size, we based our analysis on the study of Endres et al. (2017). They uncovered a large beneficial effect of an elaboration prompt on higher learning outcomes (*d*

= 0.62). Based on our power analysis ($d = 0.62$; $\alpha = 0.05$; power = .80), the necessary sample size was $N = 66$ (G*Power 3.1.9.4; Faul, Erdfelder, Buchner, & Lang, 2009).

3.2. Participants and design

In this study, we tested an initial sample of $N = 67$ participants. We had to exclude three participants due to technical problems. The remaining 64 participants were mainly university students studying psychology. The majority of the participants were female (70%) and ranged from 18 to 48 years of age ($M_{\text{age}} = 24.05$; $SD_{\text{age}} = 6.05$). The design of our immersive VR-study was a between-subject design: Participants were randomly assigned to one of the two experimental conditions based on a computer-generated randomization list (RandList 1.2; Ellwanger & Luedtke, 2001) with ($n = 31$) or without ($n = 33$) an elaboration prompt. As dependent variables, we measured the learning outcome on the three levels: knowledge, comprehension, and application. Additionally, accuracy scores based on the mental model structure were measured by the SAK, both before and after learning. As covariates, we considered learners' aptitudes, namely, their prior knowledge (functional and structural) and verbal ability. For further exploration of the VR learning experience, we assessed learners' perceived presence in VR and attention distribution. Our analysis was based on head-tracking data. This brief measurement of attention was used to gain insights into attention distribution and interaction with the VRLE.

3.3. Learning material and the elaboration prompt

The *auditory learning unit* consisted of two main topics: On the one hand, it contained details on the robot, for instance, its technical specification; on the other hand, rather abstract content such as basic kinematic definitions were included. To foster learners' engagement in the topic, we chose the robot as the narrator ('Hello, I am a service robot for home use ...'). The robot described its height, weight and the potential use of the platform on its back. Additionally, it provided more details about the general distinction between industrial and service robots. The robot gave insights into its main function, which was cleaning the floor. It described the necessary technical components to conduct its cleaning task. For instance, it described its sensors and safety protocols ('First, I capture my environment through various sensors and a camera ...'). This explanation also included details about data processing, its fusion and its transformation into appropriate reactions. Furthermore, the robot provided information about its general moving trajectory and its relation to the more abstract concept of kinematic ('I can move in a straight line, which is also described by the term ...'). Next, the robot provided more details about its locomotive system and the underlying control circuitry. For instance, it described the particular controller (PID controller) and its connection to movements and driving speed. At the end of the learning unit, the robot explained differences in local and global navigation and how this was related to its cleaning task ('... in global navigation, I have a more comprehensive environment model and try to reach my cleaning goal without driving into unnecessary dead ends ...').

While listening to the robot, the learners saw the robot as a VR animation, allowing them to visually perceive some aspects of the learning content (see Fig. 1). This animation contained information on the robot, its component, and its moving trajectory, and its purpose was to stimulate the learners to elaborate on the learning content deeply. Hence, the learning unit focused on different types of knowledge. Both, the robot's narration and VR animation of the robot contributed to the knowledge acquisition. For instance, included facts (e.g. height and weight of the robot) reflected the learning outcome's knowledge level (Bloom, 1956). The distinction between related concepts reflected the comprehension level (e.g. Which aspects are taken into account for global instead of local navigation?). The application level included a more general understanding of different components (e.g. What happens if the distance between the current pulses in the drive is shortened?).

The animation was a VR simulation programmed with Unity (version 2019.2.0f1). The learners were able to choose their exploration perspective and move in the VR. By using the controls, they were able to interact with the surrounding environment. To get to know the learning environment, learners had to perform a short sorting task of dishes in the kitchen environment. However, the learners were not able to manipulate the robot. A 3D model of the humanoid robot REEM from the manufacturer PalRobotics was used to display the virtual robot. REEM was placed in a kitchen environment and automatically entered the room. The task of the learners



Fig. 1. Virtual reality learning environment displaying the service robot.

was to listen to the robot and observe its behavior. Because the study contained a second and separate part investigating human–robot interaction, the robot model was slightly adapted to be slightly larger than its real counterpart (1.75 m instead of 1.6 m). In addition, a mouth was added to the robot's face to make it appear friendlier in the VR. The VRLE was displayed by using an *HTC Vive Pro*. Previous findings outline that stereoscopic view and wide fields of view of visual displays are essential to ensure the participants feel immersed in the VR environment (Cummings & Bailenson, 2016).

To activate learners, we included a prompt in our study before learners entered the VRLE. Prompts might trigger different cognitive learning processes by inducing generative learning strategies (Fiorella & Mayer, 2016; Bannert, 2009). The chosen prompt was developed based on theoretical considerations and adapted based on prompts that were used in previous publications (Bannert & Reimann, 2012; Berthold et al., 2009; Endres et al., 2017):

'You will now get to know the robot better. You will also learn some details about robotics. Pay particular attention to concepts and terms that are new to you and how they are related to each other. Imagine that you have to explain them to someone else who has no prior knowledge in this field.'

The developed prompt included different elements, that aimed at stimulating the elaboration of the learning content. First, we aimed to activate learners to recall the most important keywords and definitions with the chosen elaboration prompt. For that purpose, we included an explanatory component (Bannert, 2009). The publication of Berthold et al. (2009) outlines the effectiveness of including such an explanatory component into a prompt on both, procedural knowledge and conceptual understanding. Based on this finding, we asked our learners to prepare a peer-explanation at the end of the learning unit. For the peer-explanation, learners are required to recall the most important concepts and definitions and hence this might trigger learning outcome on the knowledge level (Bloom, 1956). As claimed before, our aim was not simply to induce superficial learning but to stimulate in-depth processing of learning content. Hence, we included a second aspect into the developed prompt: Learners received a hint, that the presented concepts are also related in a meaningful way. This part of the prompt was based on the study of Endres et al. (2017). They used one elaboration prompt to stimulate learners to think of potential connections or conflicts of the new information from the learning unit and their pre-existing knowledge. Knowing about meaningful connections and understanding the relationships between concepts, or deduce potential conflicts, reflects learning outcome on the comprehension level (Bloom, 1956). Hence, by including such a component into our elaboration prompt, learners were stimulated to go beyond superficial processing and to process the learning content semantically (Schnotz & Bannert, 2003). The third aspect of our developed prompt was that the peer-explanation does not only require rephrasing the basic concepts but is challenging as the peer does not have any prior knowledge in the field of robotics. Hence, learners can not simply rely on repeating the phrases of the learning unit but need to translate the learning content in their own words. This requires processing the learning content semantically and for instance search for synonyms in their mental network of prior knowledge or contrasting the new concepts to concepts that they already know (Schnotz & Bannert, 2003; Endres et al., 2017). This third aspect of our elaboration prompt reflects learning outcome on the application level (Bloom, 1956). Furthermore, the fact that learners are triggered to build meaningful connections, which requires translating and integrating the new concepts into their prior knowledge network, should stimulate them to build a more accurate structural knowledge network (Kraiger et al., 1993; Bannert & Reimann, 2012).

3.4. Questionnaires

In a short *online questionnaire*, participants were asked for their gender, age, educational level, the field of study, experiences with VR head-mounted displays and applications, prior contact with robots, and whether they own a specific type of robot. To measure verbal ability, we used the subcategory of the basic module A (similarities) of the IST2000-R (Liepmann, Beauducel, Brocke, & Amthauer, 2007). This test comprised 20 items (items 41–60) for which participants were instructed to mark two corresponding words in a list of six words. The decision should have been based on an obvious generic term suitable for two words of this list only. For instance, for the fictitious list 'fork, butter, necklace, book, biscuit, and cigarette' the correct answer is 'butter and biscuit' because both are food products.

The *pre-test questionnaire* included a measurement of prior knowledge to control different domain-specific prior knowledge levels between the experimental groups. This test aimed to measure domain-specific knowledge in science with a particular focus on relevant technical aspects related to robotics and further questions related to robots' use in real-life settings. The test comprised seven open questions (e.g., 'What is the difference between control and regulation in the field of robotics?' or 'Name three different types of sensors used in the field of robotics'). To determine the rigor of the chosen questions, we determined the inter-rater reliability, which revealed a very high consistency between the two raters ($r = 0.93$, $CI = 0.89$ – 0.96). To further explore the structural pre-existing mental model of robotics' relevant concepts, we used the SAK technique (Trumpower & Vanapalli, 2016). SAK is based on a pairwise rating of relevant concepts for their connectedness without considering their relation's causality or direction. For each pair of concepts, learners had to rate the level of connectedness on a 7-point Likert scale. In the literature, many concepts ranging between 6 and 20 have been commended (Wouters et al., 2011). Based on these expert ratings, we selected the following 6 relevant concepts of robotics: 0: PID controller, 1: transition, 2: kinematics, 3: actuator technology, 4: sensor technology, and 5: environment model. Thus, overall, learners made 36 ratings (for further details see appendix A). As recommended in literature, we referred to experts to find appropriate SAK rating concepts (Riemer & Schrader, 2016). Additionally, we ensured that rating all concepts as highly connected did not result in a higher accuracy value by choosing the concepts respectively. Two researchers of learning and instruction that were experts at extracting relevant concepts from learning materials were recruited to choose appropriate concepts for the given learning material. They first analyzed the given learning content for potential concepts. Again, an expert of research on robotics and computer science

provided feedback on the different concepts.

Furthermore, we developed a *post-test* for measuring learning outcomes on a functional level. The post-test comprised 12 open questions measuring the different learning outcomes: knowledge, comprehension, and application. To measure a learning outcome on the knowledge level, 4 recall questions were developed (e.g., ‘Which term is appropriate to describe that the robot is heading straight toward you?’). For the comprehension level, 4 questions were included (e.g., ‘What can a service robot do in contrast to an industrial robot?’), and for application, we developed 4 questions (e.g., ‘What happens if the distance between the current pulses in the actuator is shortened?’). Again, we determined the inter-rater reliability, which revealed a very high consistency between the two raters ($r = 0.88$, $CI = 0.81-0.93$).

As we used a head-mounted display enabling the learners to perceive the VRLE in 3 D and stereoscopic, we claimed our VRLE to be immersive. Hence, we included a manipulation check on how real the participants experienced the VR and their subjective presence. We used the Technology Usage Inventory (TUI) developed by Kothgassner et al. (2012). The scale comprised four items rated on a 7-point Likert scale (e.g., ‘Through the virtual simulation, I had the feeling of really experiencing the situation.’). Internal consistency was high, with Cronbach’s $\alpha = 0.84$ ($CI = 0.82-0.85$).

3.5. Procedure

This study started with an *online questionnaire*, which included demographic questions and assessed learners’ verbal ability. The questionnaire was filled in at home by the participants before the testing session. At the beginning of the testing session, which was conducted in a laboratory at the university, each participant was informed of the study’s procedure and signed an informed consent form before participating. The participants were informed that their data was handled confidentially and that they could withdraw their data or participation at any point in the study without having any disadvantage. Afterward, the participants filled out the *pre-test questionnaire*, namely, questions to assess their prior knowledge and the structure of their pre-existing mental model (SAK). This step was followed by a brief introduction to the VR equipment. Depending on the experimental condition, the participant received the elaboration prompt (see Fig. 2). To ensure that the participants understood and remembered the elaboration prompt, they had to repeat it in their own words shortly. After this, a short exploration phase in the VRLE was included. The participants conducted a short introduction task including sorting dishes to know the virtual world better as recommended in literature (e.g. Wu et al., 2020). This lasted about 2 min. Then the learning unit started as the robot entered the kitchen. After the learning unit, the participants filled in the *post-test questionnaire*, including the learning performance test, the post rating for the SAK, and the subjective rating or presence. Overall, the duration of the study was 2 hours. An additional part of investigating human-robot interaction was conducted after the described study and used one of the 2 hours.

3.6. Data preparation

All questionnaires were presented by using the online survey tool *Unipark*. Data preparation was performed by using R 3.5.1 and RStudio 1.1.463. To analyze the effect of the elaboration prompt, the two experimental groups were dummy coded (0 = without, 1 = with prompt), and for calculating models, variables were z-transformed when necessary. We measured the learners’ SAK before (pre) and after (post) the learning session. To determine the accuracy of ratings, four experts rated the relatedness of the concepts. The average similarity score between individual experts’ models was very high (0.91; Acton et al., 1994; Wouters et al., 2011). To analyze the accuracy values, we used the software *Jpathfinder* and *python* code to prepare the data. The averaged relatedness ratings of the experts (median) provided the reference standard. The software *Jpathfinder* and its scaling algorithm were used to derive similarity measures of structural knowledge quality by comparing the expert network to the learners’ individual network. The learners’ accuracy scores, ranging from 0 to 1, were calculated based on the calculations. Higher values reflected a higher accuracy of the network than the expert network (for further details see appendix A). To explore the data quality of the accuracy ratings, these were analyzed for

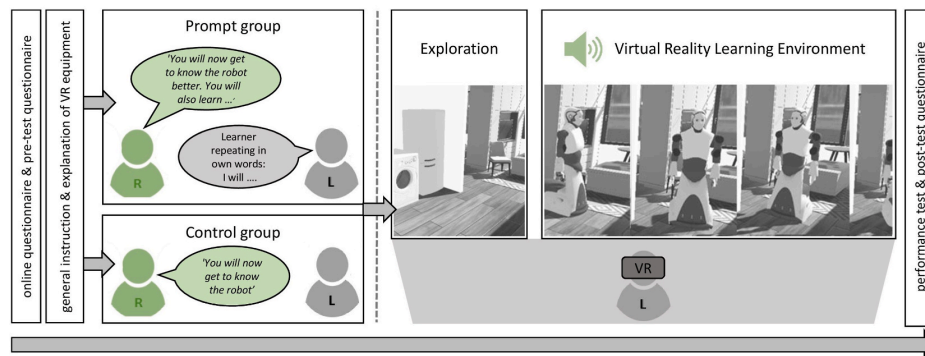


Fig. 2. Procedure for both experimental groups, displaying researcher (R) and learner (L).

fixed patterns and outliers. One participant was excluded because the SAK ratings were an outlier.

4. Results

4.1. Descriptive results

Learners' domain-specific prior knowledge was on a medium level in both experimental conditions (Table 1). The learning outcome was on a medium level, both overall and on the knowledge and comprehension level. The learning outcome values for application were low (Table 1).

To ensure that the two experimental groups did not differ significantly in their relevant characteristics, a MANOVA was conducted with age, prior knowledge, verbal ability, presence, and pre-accuracy for the rating of the SAK as dependent variables. No significant differences between the groups could be found ($F(5,55) = 0.69, p = .694$). Moreover, the groups did not differ in their gender distribution ($\chi^2(1, N = 64) = 0.01, p > .050$). As previously mentioned, learners' prior knowledge and verbal ability play an important role in learning outcome. Therefore, these factors were included as covariates in the analyses. Because of missing data, six cases were excluded in these analyses when the verbal ability was included. Based on a conducted Shapiro-Wilk test, the multivariate normal distribution can be assumed for the variables learning outcome, prior knowledge, and accuracy (pre, post) for each experimental subgroup ($p > .056$), and the variances can be classified as homogenous based on Bartlett's test ($p > .367$).

4.2. Effect of the prompt on learning outcome

To test our first hypothesis, we analyzed the effect of the elaboration prompt on the functional level by differentiating the learning outcome on the knowledge, comprehension, and application level (Bloom, 1956). We expected a significantly higher learning outcome in the prompt group than in the control group on all three levels of the learning outcome (H1a).

The prompt group's means were higher for overall learning outcome, knowledge, and application compared to the control group, as can be seen in Table 2. For comprehension, the group without prompt showed a descriptively higher learning outcome. To test the prompt's main effect on overall learning outcome, we conducted an ANCOVA including verbal ability and prior knowledge as covariates. We found no significant effect of the elaboration prompt on overall learning outcome ($F(1,55) = 0.19, p = .661$). Furthermore, to analyze the elaboration prompt effects on each of the three levels of learning outcome, we conducted a MANCOVA ($F(3,55) = 1.14, p = .172, \eta^2_{\text{partial}} = .059$). When controlling for variance because of verbal ability and prior knowledge the results show a significant, beneficial effect of the prompt on learning outcome on the knowledge level, only ($p = .036$; see Table 3). No significant effects were found on comprehension and application level. Additionally, we expected the largest beneficial effect on the higher learning outcome levels (comprehension and application; H1b). The descriptive pattern is already not in line with our previous assumption because for comprehension the prompt had no positive effect. Therefore, this hypothesis is rejected.

4.3. Effect of the prompt on structural accuracy

To uncover the elaboration prompt's effects on the structural level, we measured the structural accuracy deduced from the SAK ratings after the learning session. In our second hypothesis, we expected that learners in the prompt group had a significantly higher post-accuracy score than the control group after accounting for pre-accuracy. In detail, we expected that the beneficial effect of the elaboration prompt depends on the pre-accuracy score (H2). Looking at the visualization of the post networks of the different groups, the configuration of experts, prompt group and control group seem to differ substantially based on post-accuracy ratings. The prompt group ratings indicate that more connections and closer relationships between the individual concepts were assumed than in the expert or control group (see Fig. 3). For instance, concept 0 (PID controller) and 5 (environment model) were not connected based on experts rating while concept 2 (kinematics) was rated as highly connected with 3 (actuator technology). In contrast, learners in the control group rated the concepts 0 and 5 to be closer connected compared to 2 and 3. This results in lower post-accuracy ratings compared to the learners in the prompt group who rated 2 and 3 as closer connected compared to 0 and 5. To be able to interpret these differences, we calculated accuracy scores for each experimental group before and after learning in VR (see appendix A).

The accuracy means compared to the experts' ratings were in both groups higher after the learning unit than the baseline measurement (see Table 4). Descriptively, the highest post-accuracy mean was found for the prompt group.

Overall, there was a significant increase from pre- to post-accuracy ($t(62) = -3.31, p < .001, r = 0.79$). As our research question

Table 1
Means and standard deviations of the experimental conditions.

	With prompt $n = 31$ $M (SD)$	Without prompt $n = 33$ $M (SD)$
Prior knowledge (%)	51.44 (12.40)	47.21 (14.01)
Verbal ability (%)	58.01 (13.07)	53.00 (15.00)
Pre-accuracy (SAK, $Max = 1$)	0.42 (0.12)	0.40 (0.10)
Age (years)	24.10 (6.62)	24.21 (5.29)
Male (%)	30.30	38.71

Table 2
Means and standard deviations of learning outcome depending on the experimental condition.

	With prompt <i>M (SD)</i>	Without prompt <i>M (SD)</i>
Overall learning outcome (%)	39.93 (12.62)	37.54 (12.78)
Knowledge (%)	66.41 (22.55)	58.87 (23.76)
Comprehension (%)	32.39 (16.61)	34.54 (16.66)
Application (%)	21.09 (15.48)	19.22 (17.00)

Table 3
Results of the ANCOVA depending on the level of learning outcome.

	<i>F</i> (1,55)	<i>p</i>	η^2_{partial}
<i>Knowledge</i>			
Prompt	3.42	.036*	.042
Verbal ability	0.29	.296	.013
Prior knowledge	4.49	.019*	.075
<i>Comprehension</i>			
Prompt	0.03	.432	.006
Verbal ability	1.42	.119	.018
Prior knowledge	1.36	.124	.024
<i>Application</i>			
Prompt	0.03	.436	.007
Verbal ability	8.82	.002**	.120
Prior knowledge	1.97	.083	.035

Note. ** $p < .01$, * $p < .05$.

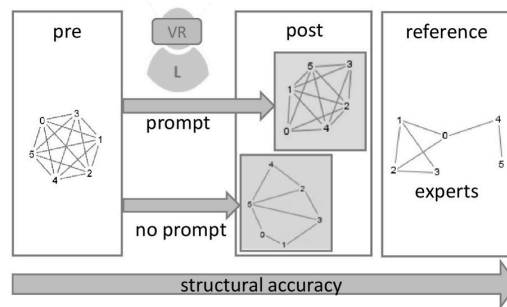


Fig. 3. Knowledge networks, before and after learning in VR contrasted to the experts' knowledge network; from left to right with increasing accuracy scores.

aimed at answering the question of how the effect of the prompt on post-accuracy depends on the pre-accuracy score, we chose an ANCOVA including an interaction term as recommended by Wan (2019) for pre-accuracy and experimental condition (prompt vs. no prompt) as well as the covariate verbal ability (see Table 5).

We found the expected significant interaction of the experimental condition and pre-accuracy on post-accuracy ($F(1,52) = 3.11, p = .042, \eta^2_{\text{partial}} = .056$). Therefore, the effect of the elaboration prompt on post-accuracy depends on the pre-accuracy. For better understanding this interaction, we visualized the effects of pre-accuracy by dividing learners into hypothetical groups with three distinct levels of pre-accuracy in Fig. 4.

Learners with low pre-accuracy benefit from the elaboration prompt compared to the group without prompt. For learners with medium and high pre-accuracy, this beneficial effect was not found.

Table 4
Means and standard deviations of learning outcome depending on the experimental condition.

	With prompt <i>M (SD)</i>	Without prompt <i>M (SD)</i>
Pre-accuracy	0.42 (0.12)	0.40 (0.10)
Post-accuracy	0.51 (0.13)	0.47 (0.14)

Table 5

Results of the ANCOVA with the experimental condition, verbal ability, and pre-accuracy on post-accuracy.

	$F(1,52)$	p	η^2_{partial}
Prompt	0.77	.192	.017
Verbal ability	2.05	.079	.027
Pre-accuracy	0.33	.285	.006
Prompt*pre-accuracy	3.11	.042*	.056

Note. * $p < .05$.

4.4. Relationship of learning outcome based on performance and structural accuracy

Both dependent variables, learning outcome and SAK, aimed to measure learners' mental models by considering different perspectives. In our third hypothesis, we expected a significant positive correlation of the learning outcome with the application level and post-accuracy values (H3). In line with our expectations, application and post-accuracy showed a moderate, significant, positive correlation ($r = 0.36$, $p < .010$; Table 6). Neither knowledge nor comprehension level showed a significant correlation with post-accuracy scores.

4.5. Effect of the prompt on attention distribution

To gain further insights into the VR experience of the learners, we measured their perceived presence and explored their behavior via head-tracking. The perceived presence was rather high in both groups ($M_{\text{prompt}} = 5.52$, $SD_{\text{prompt}} = 0.87$, $M_{\text{noprompt}} = 5.14$, $SD_{\text{noprompt}} = 1.10$). To ensure that learners focused on their learning task, we recorded their field of view while learning in VR. To quantify whether the learners were distracted by the VR environment, we analyzed the time when the robot was in their field of view completely, indicating that it was focused on by the learner. Due to technical problems, while recording the data, only videos of $n = 42$ learners were available. Based on the descriptive means, learners with prompt focused on the robot longer ($M_{EG} = 88\%$, $SD_{EG} = 18$) than the learners without prompt ($M_{CG} = 79\%$, $SD_{CG} = 35$). This difference was not significant ($t(40) = -1.2$, $p = .117$, $d = 0.32$).

5. Discussion

In this study, we investigated the effects of an elaboration prompt that aimed to activate learners to process the learning content deeply and thus semantically when learning in VR. The chosen prompt included elements to foster the elaboration of the learning content and pointed out that the concepts should not only be considered in isolation but that the deduction of possible connections between the concepts were also part of the learning process. Additionally, the prompt contained an explanatory component as learners were asked to prepare a peer explanation at the end of the learning session. To examine the prompt's effect on the resulting mental model, we used a multi-method approach: one functional approach with a differentiated learning outcome measurement (knowledge, comprehension, and application) and one structural approach by using the SAK technique. By this, we aimed to gain further insights into the developmental process of mental models in VRLEs.

5.1. Impact of the elaboration prompt on the functional and structural level

In our first hypothesis, we expected a beneficial effect of the prompt on the learning outcome, particularly for the higher levels of learning outcome: comprehension and application. In contrast to this expectation, the prompt led to beneficial effects on the knowledge level, only. Hence, the used elaboration prompt did not foster in-depth and semantic processing. As comprehension and

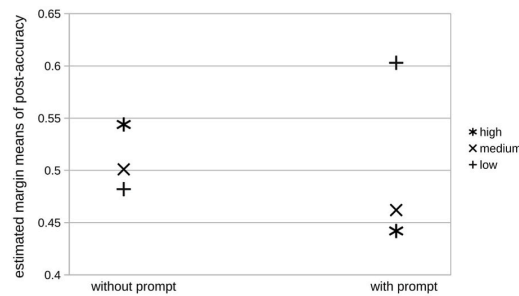


Fig. 4. Post-accuracy depending on experimental group and pre-accuracy. Note: Pre-accuracy defined in low $< \text{mean} - 1 \text{ SD}$: $n = 4$ with and $n = 5$ without prompt; medium = $\text{mean} \pm 1 \text{ SD}$: $n = 21$ with and $n = 23$ without prompt; high $> \text{mean} + 1 \text{ SD}$: $n = 6$ with and $n = 4$ without prompt; the covariate verbal ability is constant with $\beta_{\text{verbal}} = 0.56$.

Table 6
Relationship between learning outcome and structural assessment of knowledge (SAK).

	SAK _r	<i>p</i>
Knowledge	.19	.141
Comprehension	.09	.460
Application	.36	.004**

Note. ***p* < .01.

application scores were low in both experimental groups, overall, very little semantic processing occurred. Thus, based on the present findings, learners were unable to build meaningful connections between the different concepts and were therefore unable to predict the consequences of manipulating one component onto another. Other studies have also described limited cognitive information processing when learning in VRLEs (e.g. Makransky et al., 2020; Makransky, Terkildsen, & Mayer, 2019; Wu et al., 2020). Whether VRLEs were effective learning environments depended largely on the boundary conditions such as the subject or the chosen sample (Wu et al., 2020). In our study, the limited semantic processing might be because the learning unit included many technical and abstract concepts. Consequently, the learners might have had insufficient resources to develop a causal mental model that would answer the questions on these levels (Bloom, 1956; Mayer, Mathias, & Wetzell, 2002). Hence, VRLE settings do not per se guarantee adequate semantic processing even when an elaboration prompt stimulated learners to process the content in-depth.

What about the structural level? Does the prompt enable the learners to have a basic understanding of the relationships between the concepts? We addressed this question in our second hypothesis. As expected, we found a significant interaction of prompt and pre-accuracy on structural accuracy after learning. Hence, learners' structural accuracy depended on both the quality of their pre-existing knowledge network and whether or not they received the elaboration prompt. Learners whose pre-accuracy was in the low or medium range, benefit from the elaboration prompt. An expertise reversal effect was found for learners with higher pre-accuracy when they were prompted. In line with prior findings, supporting learners with higher prior knowledge might interfere with their strategies or their pre-existing knowledge network (Kalyuga, 2009).

5.2. Coherence of the functional and structural approach

In our third hypothesis, we expected a positive relationship between the application level and structural accuracy. As hypothesized, we found a moderate positive correlation between application and structural accuracy. In line with our expectation, no significant correlation was found for knowledge or comprehension. Answering application questions includes deducing correct causal or directional connections of the components (Bloom, 1956; Schnotz & Bannert, 2003). By contrast, it is sufficient for the structural accuracy to recognize these concepts' pure connectedness (Kraiger et al., 1993). Therefore, both approaches require a more global understanding of the relevant concepts' connection in a more or less sophisticated mental model. Although the concepts have this in concordance, they differ significantly in the different prerequisites for the specificity of the different components' connections. The two first hypotheses' results reflected this: The elaboration prompt had beneficial effects on the knowledge level and on structural accuracy. As outlined before, high structural accuracy after learning reflects information on the relatedness of the different concepts without considering the details on the cause-and-effect relations. This information processing is also reflected in the IMTPC as a shortcut between the text surface representation and the analog mental model (Schnotz & Bannert, 2003). Our findings indicate that this shortcut was stimulated by the current learning setting and allows to deduce relations on a superficial level only. For instance, the information presented in the same sentence might be stored as more closely related than information with a greater syntactical distance. Based on our findings, the processing of this relational information was stimulated successfully by the elaboration prompt, but only by the shortcut because the elaboration prompt did not increase performance on higher learning outcomes and thus no semantic processing occurred.

5.3. Limitations and recommendations for further research

This study's results indicate a limited effect of the elaboration prompt on semantic processing of the learning content; thus, VRLEs and the learning processes in such environments must be further explored. In our study, we used one prompt in the instruction before entering the VRLE. As outlined earlier, different concepts of how and when to support learners by prompting exist. The chosen prompt did simply outline the necessity to learn the concepts and their connectedness by semantic processing but did not include further strategic support for cognitive or metacognitive strategies. For instance, to support cognitive elaboration, learners could be asked to connect the learned concepts to their lives or think of further examples (e.g., Endres et al., 2017). However, when learning abstract concepts combined with very low prior knowledge, finding adequate strategy activators remains a challenge. Even when learners are motivated to use the proposed strategies, they might not have the required skills or the provided strategic activators might interfere with their existing strategies. Further possibilities to support learners include using metacognitive prompts, which might activate and support learners' self-regulation processes and their metacognitive strategy use. By this, learners might be supported on semantic processing of the learning content (Gentner & Seufert, 2020). Additionally, studies have outlined the benefits of using a combination of cognitive and metacognitive prompts on learning outcomes (Berthold et al., 2007).

In our VRLE, we aimed to display the robot's natural behavior, for instance, its moving trajectory. To avoid the robot's intimidating effect, we implemented a security protocol that ensured a minimal distance of 1.5 m between the learner and the robot. Therefore, learners could manipulate the surrounding objects in the room and choose their position to observe the robot individually but were not able to directly manipulate the robot. The possibility to compose and decompose the different parts of the robot flexibly might foster a deeper understanding of the interplay of its technical components (Kamarainen et al., 2018). Hence, direct manipulation of the robot might be implemented in future studies.

Furthermore, future studies might gain further insight into the perceived load while learning in VR. The chosen learning content in this study was complex and rather abstract. Thus, a high overall cognitive load could have occurred (Sweller, 2011). Although the learners were activated to elaborate on the learning content, they might not have had the required cognitive resources. To provide insights into the perceived cognitive load while learning, further research should consider a differentiated cognitive load measurement (Klepsch, Schmitz, & Seufert, 2017). Additionally, the perceptual load could be considered for analyzing cognitive learning processes in VRLEs (Huang et al., 2019).

Moreover, because the score on comprehension levels and the application level were low, the respective questions should be revised critically. Different questions with more variance in their difficulty might offer further insights in the resulting mental model. Furthermore, SAK is described as a valid method to gain insight into the structural components of the relevant mental models (Kraiger et al., 1993; Trumpower & Vanapalli, 2016); nevertheless, its limitations must be considered. The process of choosing the concepts must be discussed. These concepts have been chosen based on expert ratings, a standard method to deduce relevant concepts. However, choosing these concepts included to some degree arbitrary decisions. Different concepts might result in different ratings and effects. Therefore, further research might include different concepts or different subsets of concepts to investigate whether our findings are replicable. One additional challenge using the SAK is to choose the concepts in a way that rating concepts as highly connected do not automatically result in high accuracy values. In our study, we ensured that this effect did not bias the results as based on theoretical consideration, developing a knowledge network reflects building a more organized and elaborated structure. Hence, learners need to be able to differentiate what is and what is not closely connected and not just simply describe a structure that includes more connections and is, therefore, more complex (Trumpower & Vanapalli, 2016; Haiyue & Yoong, 2010; Ifenthaler, Masduki, & Seel, 2011). To assure the expert ratings were of good quality, we analyzed the consistency of the expert rating, which was very high compared to that of other studies and more extensive than the required value (Riemer & Schrader, 2016). For structural accuracy, learners only needed to rate the given concepts 'passively' and were not asked to name or verbalize the concepts. This gap between the passive knowledge of concepts and the active retrieval of information from the mental model and the successful use has often been described. Publications have outlined the importance of using adequate learning strategies to overcome this gap and thus this might be considered when designing future studies (Fan, 2000).

Moreover, different approaches to investigate the structure and the developmental process of mental models might be used. For instance, the *network elaboration technique* of Eckert (2000) might deliver further insights into the semantic and sub-semantic processing of learning content. When using this technique, learners create their individual concept maps and specify the connections between the included concepts. Extending this essential measurement of the meaningful connection, other techniques that include a more detailed description of these meaningful connections would be used in the *model of explanatory coherence* approach (e.g., Read & Marcus-Newhall, 1993). Participants are asked to describe relevant scenarios and explanatory hypotheses related to the learning content. This enables the researcher to uncover underlying principles and gain deeper insights into learners' mental models (Thagard, 1989). However, these methods are very effortful and require very motivated learners to be able to interpret the results.

By assessing individual learners' aptitudes, other concepts such as motivation and the prompt's use could be investigated (Nückles, Roelle, Glogger-Frey, Waldeyer, & Renkl, 2020). Additionally, further cognitive abilities such as logical abilities or field-independence might explain individual differences while learning in VRLEs. Finally, our chosen sample size was based on our a priori power analysis following the recommendations of Faul et al. (2009). However, as our effect size was relatively small, we recommend a higher sample size in future studies and to conduct a power analysis including a higher test power. To gain further insights into the interaction of prompting and pre-accuracy on post-accuracy, we used three hypothetical groups. This approach involved an extreme group comparison and therefore contained few learners in some groups. Therefore, the described findings need to be further validated by higher sample sizes in future studies (Preacher, Rucker, Mac Callum, & Nicewander, 2005).

5.4. Conclusion

Before learning in VR, prompting the learners to elaborate on the learning content showed beneficial effects. Although, this manipulation of including a short elaboration prompt was a very brief intervention, it showed a substantial effect on the knowledge level and the structural level. Hence, based on our findings, elaboration prompts are recommended for future VRLEs, particularly for learners with lower to medium prior knowledge levels. Although the chosen elaboration prompt showed beneficial effects on the recall or knowledge level only, we outlined this level as a critical precondition for further semantic processing and a deeper understanding of the learning content. Using the multi-method approach to measure mental model enables further deduction of details on what learners processed semantically and how they organized this knowledge structurally. In our study, learners were not able to answer the application questions correctly. Hence, they could not build correct meaningful connections, even when being activated by the elaboration prompt. However, they were supported by the prompt in building a structural mental model with unspecific connections reflected by a less sophisticated mental model. Many challenges to developing adequate instructional designs and supportive elements for VRLEs remain for further research. This study outlines the importance of investigating the boundary conditions of learning in VRLEs to compare their effectiveness to conventional learning settings.

Author contributions

AV developed the design of the learning study as well as the auditory learning material (Conceptualization & Methodology). PH and FB created the VR learning environment (Software; Visualization). AV and FB led the data collection for the study. TS and MB feed backed the study design (Supervision). AV analyzed and interpreted the data. AV drafted the work (writing – original draft), which was revised critically by FB, PH, MB and TS (review). All authors provided approval of the final submitted version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethical statement

The study was carried out in accordance with the Declaration of Helsinki. Ethical approval was received from the ethical committee of Ulm University (No. 378/19). The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Declaration of competing interest

None: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This study was funded by the Federal Ministry of Education and Research Germany (grant number: 16OH22032). We like to thank our students and the consulted experts involved this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compedu.2021.104235>.

References

- Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, 86(2), 303. <https://doi.org/10.1037/0022-0663.86.2.303>
- Baceviciute, S., Mottelson, A., Terkildsen, T., & Makransky, G. (2020). Investigating representation of text and audio in educational VR using learning outcomes and EEG. In R. Bernhaupt, F. F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, et al. (Eds.), *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376872>.
- Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie*, 23(2), 139–145. <https://doi.org/10.1024/1010-0652.23.2.139>
- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. <https://doi.org/10.1007/s11251-011-9167-4>
- Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, 37(4), 345–363. <https://doi.org/10.1007/s11251-008-9051-z>
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577. <https://doi.org/10.1016/j.learninstruc.2007.09.007>
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals (Handbook I: Cognitive domain)*. New York: David McKay Co Inc.
- Burbules, N. C. (2006). Rethinking the virtual. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *The international handbook of virtual learning environments* (pp. 37–58). Dordrecht: Springer. Retrieved from <https://dx.doi.org/10.1007/978-1-4020-3803-7>.
- Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. *Computers & Education*, 159, 104005. <https://doi.org/10.1016/j.compedu.2020.104005>
- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2), 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating cognitive training outcomes: Validity and utility of structural Knowledge Assessment. *Journal of Business and Psychology*, 18(2), 191–206. <https://doi.org/10.1023/A:1027397031207>
- Eckert, A. (2000). Die Netzwerk-Elaborations-Technik (NET) – Ein computerunterstütztes Verfahren zur Diagnose komplexer Wissensstrukturen. In H. Mandl, F. Fischer, & Hrsg (Eds.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (pp. 137–157). Hogrefe.
- Ellwanger, U., & Luedtke, H. (2001). *RandList. DatInf GmbH Version 1.2*. <http://randomisation.eu/index.shtml>.
- Endres, T., Carpenter, S., Matrin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13–20. <https://doi.org/10.1016/j.learninstruc.2016.11.010>
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105–119. <https://doi.org/10.1177/003368820003100205>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
- Gentner, N., & Seufert, T. (2020). The double-edged interactions of prompts and self-efficacy. *Metacognition and Learning*, 15(2), 261–289. <https://doi.org/10.1007/s11409-020-09227-7>

- Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83(1), 88–96. <https://doi.org/10.1037/0022-0663.83.1.88>
- Haiyue, J., & Yoong, W. K. (2010). A network analysis of concept maps of triangle concepts. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the mathematics education research group of australasia*. Fremantle, Australia: MERGA.
- Hong, E., & O'Neil, H. F. (1992). Instructional strategies to help learners build relevant mental models in inferential statistics. *Journal of Educational Psychology*, 84(2), 150–159. <https://doi.org/10.1037/0022-0663.84.2.150>
- Howard, M. C., & Lee, J. (2020). Pre-training interventions to counteract seductive details in virtual reality training programs. *Human Resource Development Quarterly*, 31(1), 13–29. <https://doi.org/10.1002/hrdq.21378>
- Huang, K. T., Ball, C., Francis, J., Ratan, R., Boumis, J., & Fordham, J. (2019). Augmented versus virtual reality in education: An exploratory study examining science knowledge retention when using augmented reality/virtual reality mobile applications. *Cyberpsychology, Behavior, and Social Networking*, 22(2), 105–110. <https://doi.org/10.1089/cyber.2018.0150>
- Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science*, 39(1), 41–61. <https://doi.org/10.1007/s11251-009-9097-6>
- Johnson-Laird, P. N. (1981). Mental models in cognitive science. *Cognitive Science*, 4(1), 71–115. [https://doi.org/10.1016/S0364-0213\(81\)80005-5](https://doi.org/10.1016/S0364-0213(81)80005-5)
- Jonassen, D. H. (1995). Operationalizing mental models: Strategies for assessing mental models to support meaningful learning and design-supportive learning environments. In *Proceedings of CSCI '95: The first international conference on computer support for collaborative learning* (pp. 182–186). Lawrence Erlbaum Associates.
- Kalyuga, S. (Ed.). (2009). *Managing cognitive load in adaptive multimedia learning*. Hershey, Pa: Information Science Reference. <https://doi.org/10.4018/978-1-60566-048-6>
- Kamarainen, A. M., Thompson, M., Metcalfe, S. J., Grotzer, T. A., Tutwiler, M. S., & Dede, C. (2018, June). Prompting connections between content and context: Blending immersive virtual environments and augmented reality for environmental science learning. In *International conference on immersive learning* (pp. 36–54). Cham: Springer.
- Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology*, 51(6), 2115–2138. <https://doi.org/10.1111/bjelt.13029>
- Kothgassner, O. D., Felhofer, A., Hauk, N., Kastenhofer, E., Gomm, J., & Kryspin-Exner, I. (2012). TUI - technology usage inventory. https://www.fhg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf
- 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(2), 311–328. <https://doi.org/10.1037//0021-9010.78.2.311>
- Lehmann, J. A., & Seufert, T. (2018). Can music foster learning—Effects of different text modalities on learning and information retrieval. *Frontiers in Psychology*, 8, 2305. <https://doi.org/10.3389/fpsyg.2017.02305>
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-struktur-test 2000 R (2. Aufl.)*. Hogrefe.
- Loibl, K., & Leuders, T. (2019). How to make failure productive: Fostering learning from errors through elaboration prompts. *Learning and Instruction*, 62, 1–10. <https://doi.org/10.1016/j.learninstruc.2019.03.002>
- Makransky, G., Andreassen, N. K., Baccviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000473>. Advance online publication.
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>
- Mayer, R. (2014). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>
- Mayer, R. E., Mathias, A., & Wetzel, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, 8(3), 147. <https://doi.org/10.1037/1076-898X.8.3.147>
- Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140, 103603. <https://doi.org/10.1016/j.compedu.2019.103603>
- Newmeyer, F. J. (1996). Linguistics. In *The Cambridge survey: Volume 2, linguistic theory: Extensions and implications*. Cambridge Univ. Press (Original work published 1988).
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, 1–38. <https://doi.org/10.1007/s10648-020-09541-1>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797. <https://doi.org/10.1037/edu0000241>
- Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, 1–16. <https://doi.org/10.1111/jcal.12482>
- Preacher, K. J., Rucker, D. D., Mac Callum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, 10(2), 178–192. <https://doi.org/10.1037/1082-989X.10.2.178>
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, 65(3), 429.
- Reigeluth, C. M. (1983). *Instructional-design theories and models: An overview of their current status*. Lawrence Erlbaum Associates.
- Riemer, V., & Schrader, C. (2016). Impacts of behavioral engagement and self-monitoring on the development of mental models through serious games: Inferences from in-game measures. *Computers in Human Behavior*, 64, 264–273. <https://doi.org/10.1016/j.chb.2016.06.057>
- Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349–363. <https://doi.org/10.1037/0033-2909.100.3.349>
- Scheiter, K., Schüler, A., & Eitel, A. (2017). Learning from multimedia: Cognitive processes and instructional support. In S. Schwan, & U. Cress (Eds.), *The psychology of digital learning: Constructing, exchanging, and acquiring knowledge with digital media* (pp. 1–19). Cham, Switzerland: Springer.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schnotz, W., Böckeler, J., & Grzonziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, 14(2), 245–265. <https://doi.org/10.1007/BF03172968>
- Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.), *The psychology of learning and motivation: Vol. 55. The psychology of learning and motivation: Cognition in education* (pp. 37–76). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12(3), 435–502.
- Trumpower, D. L., & Vanapalli, A. S. (2016). Structural assessment of knowledge as, of, and for learning. In J. M. Spector, B. B. Lockee, & M. Childress (Eds.), *Learning, design, and technology: An international compendium of theory, research, practice, and policy* (pp. 1–22). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_23-1
- Wallen, E., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research & Development*, 53(3), 59–71. <https://doi.org/10.1007/bf02504798>

- Wan, F. (2019). Analyzing pre-post randomized studies with one post-randomization score using repeated measures and ANCOVA models. *Statistical Methods in Medical Research*, 28(10–11), 2952–2974. <https://doi.org/10.1177/0962280218789972>
- White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42(1), 99–157. [https://doi.org/10.1016/0004-3702\(90\)90095-h](https://doi.org/10.1016/0004-3702(90)90095-h)
- Wouters, P., van der Spek, E., & van Oostendorp, H. (2011). Measuring learning in serious games: A case study with structural assessment. *Educational Technology Research & Development*, 59, 741–763. <https://doi.org/10.1007/s11423-010-9183-0>
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005. <https://doi.org/10.1111/bjet.13023>

Appendix A

Structural Assessment of Knowledge

In the present study, we used a multimethod approach to investigate learning outcome and mental models. Next to using a performance test to deduce functional aspects of learning outcome, we chose a structural approach. To gain further insights into the underlying mechanisms of the *Structural Assessment of Knowledge* technique (SAK), we present and define relevant concepts and provide additional information about the underlying algorithms in the following.

The SAK is a psychometric scaling method based on graph theory and researchers used this method in the context of knowledge acquisition. A basic assumption of mental models' structural measurements that declarative knowledge is relational (Kraiger, Ford, & Salas, 1993). This means that individual concepts are not considered in isolation but are integrated into a complex knowledge network when the mental model has developed accordingly (Davis, Curtis, & Tschetter, 2003). Therefore, knowledge acquisition is not completed by mere storage or semantic processing, but the accumulated knowledge must also be organized (Kraiger et al., 1993). Hence, in this context, one can also label mental models with terms like knowledge structure, cognitive maps or task schemata (Rouse & Morris, 1986; Kraiger et al., 1993). Thus, mental models also provide a way to contextualize and understand what has been learned. Furthermore, this learning content is structured and organized. Additionally, the existing knowledge network also influences the acquisition of new knowledge to a substantial degree (Kraiger et al., 1993).

Thus, learning involves understanding similarities and relationships between concepts, facts, definitions, processes and procedures (Davis et al., 2003). A structural assessment of learning involves three major steps (Wouters, van der Spek, & van Oostendorp, 2011): First, learners need to judge the concepts' relationships. Then a cognitive structure based on these judgments needs to be derived. As the last step, the resulting cognitive structures need to be evaluated by matching these with experts' knowledge structures (Goldsmith & Kraiger, 2014).

To understand this judgment to deduce cognitive structures, we contrast the terms *proximity* and *similarity*. Both terms are closely related and aim to describe the degree to which concepts of the SAK belong to each other from a psychological perspective (Schvaneveldt, 1990). The concept proximity refer to both, subjective and objective relationship between pairs of concepts. Therefore, to deduce distance between pairs of concept often relatedness (proximity) values are reported (Trumpower & Vanapalli, 2016). Higher proximity values reflect a higher rated relatedness of pairs of items. These connections between pairs of concepts include distance estimates. The term similarity is used to describe two different networks and to which degree they are similar. To be able to gain further insight into the accuracy of the learners' networks, referential networks of experts are chosen as criteria (Wouters et al., 2011; Goldsmith & Kraiger, 2014). These are aggregated and then

compared to the individual rating of the learners. This similarity networks, or as named in the present study accuracy values as calculated as follows:

To compare the network of the experts and the learners, these need to include the same concepts and paired ratings. As recommended for the given data, the median ratings were computed (Davis et al., 2003). These ratings can be graphically displayed and be briefly described. In figure A1, three networks are displayed. Network a) represents the configuration of the expert, and therefore the referential network. Networks b) and c) display the networks under different experimental conditions. While learners in the experimental group with prompt rated the concepts as closely related, the learners without prompt tend to describe a higher distance and less relatedness between them.

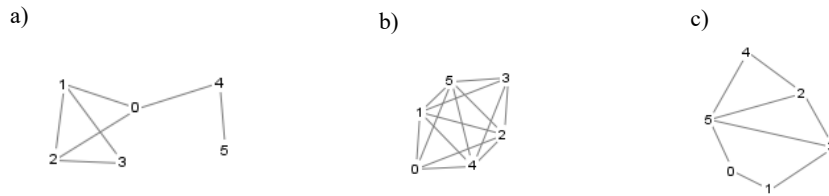


Figure A1 Network based on the post SAK ratings: a) experts, b) experimental group (with prompt), c) control group.

The generated network is based on two parameters (Schvaneveldt, 1990). The first parameter, namely the Minkowski r -metric, describes how the distance between two concepts, that are not directly linked, is computed (Kivlighan & Tibbits, 2012; Nash & Nash, 2003). This value ranges between 1 and infinity (Filiz & Trumpower, 2014). The second relevant parameter is the q parameter, which is limited by the number of available links between the different concepts (Nash & Nash, 2003). Hence, its value reflects the maximum number of links in paths in which the triangle inequalities are guaranteed to be satisfied for the resulting network (Schvaneveldt, 1990). After both parameters have been determined, two of the concepts are now assigned for linking. This happens if an indirect path's length is equal or longer than the length of the connection between these two concepts (Nash & Nash, 2003). This is a direct link between the number of links in the network and the values of r and q , when these are increased, the number of links in the network decreases (Filiz & Trumpower, 2014). In their publication, Filiz and Trumpower (2014) describe the six necessary steps starting with rating the concepts and Pathfinder algorithm in detail. After the network has been calculated, these can be compared to referential or expert networks. To be able to quantify the quality of the resulting network structure, similarity scores are used as an indicator of what we defined as structural accuracy. To deduce the similarity between two networks, the number of common links is divided by the total number of unique links in the two networks. Hence, two identical and completely similar networks will result in a structural accuracy score of 1. In contrast, two networks that share no common links will result in a structural accuracy score of 0. Hence, the results reflect the proportion of all the links in the two given networks in both networks.

Additional References

- Davis, M. A., Curtis, M. B., & Tschetter, J. D. (2003). Evaluating Cognitive Training Outcomes: Validity and Utility of Structural Knowledge Assessment. *Journal of Business and Psychology*, 18(2), 191–206. <https://doi.org/10.1023/A:1027397031207>
- Filiz, M., & Trumpower, D. (2014). Exploring the Mobile Structural Assessment Tool: Concept Maps for Learning Website. *Revista Colombiana De Estadística*, 37(2Spe), 297–317. <https://doi.org/10.15446/rce.v37n2spe.47939>
- Goldsmith, T. E., & Kraiger, K. (2014). Applications of Structural Knowledge Assessment to Training Evaluation. In J. K. Ford (Ed.), *Applied Psychology Series. Improving training effectiveness in work organizations* (1997th ed., pp. 73–94). Hoboken: Taylor and Francis.
- Kivlighan, D. M., & Tibbits, B. M. (2012). Silence is mean and other misconceptions of group counseling trainees: Identifying errors of commission and omission in trainees' knowledge structures. *Group Dynamics: Theory, Research, and Practice*, 16(1), 14–34. <https://doi.org/10.1037/a0026558>
- 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(2), 311–328. <https://doi.org/10.1037//0021-9010.78.2.311>
- Nash, J. G., & Nash, J. M. (2003). A structural representation of migraine diagnostic criteria: The experts' view. *Headache*, 43(4), 322–329. <https://doi.org/10.1046/j.1526-4610.2003.03066.x>
- Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349–363. <https://doi.org/10.1037/0033-2909.100.3.349>
- Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization*: Ablex Publishing.
- Trumpower, D. L., & Vanapalli, A. S. (2016). Structural Assessment of Knowledge as, of, and for Learning. In J. M. Spector, B. B. Lockee, & M. Childress (Eds.), *Learning, design, and technology*:

An international compendium of theory, research, practice, and policy (Vol. 86, pp. 1–22). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_23-1

Wouters, P., van der Spek, E. D., & van Oostendorp, H. (2011). Measuring learning in serious games: a case study with structural assessment. *Educational Technology Research and Development*, 59(6), 741–763. <https://doi.org/10.1007/s11423-010-9183-0>

The third paper was published in the British Journal of Educational Technology under the terms of the Creative Commons CC-BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>):

Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome. *British Journal of Educational Technology*, 00, 1–19.

<https://doi.org/10.1111/bjet.13104>

Journal Metrics: 2-Year Impact Factor: 3.241 (2020), h-index 95

Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome

Andrea Vogt¹ | Franziska Babel² | Philipp Hock² |
Martin Baumann² | Tina Seufert¹

¹Department Learning and Instruction,
Institute of Psychology and Education, Ulm
University, Ulm, Germany

²Department Human Factors, Institute of
Psychology and Education, Ulm University,
Ulm, Germany

Correspondence

Andrea Vogt, Department Learning and
Instruction, Institute of Psychology and
Education, Ulm University, Albert-Einstein-
Allee 47, Ulm 89081, Germany.
Email: andrea.vogt@uni-ulm.de

Funding information

This study was funded by the Federal
Ministry of Education and Research
Germany (grant number: 16OH22032)

Abstract

Presenting a pictorial representation followed by a textual representation supports learners to build a coherent mental model. Providing an elaboration prompt stimulates learners to process the learning content semantically. Hence, combining both approaches might result in synergetic effects as both foster mental model development, which could be particularly challenging in immersive virtual environments. In our 2×2 between-subject design study, we investigated the beneficial effects of adequate sequencing of text and picture and prompting, as well as their potential synergic effect in an immersive virtual reality (VR) learning environment with $N = 81$ learners. We presented the learning content in form of one pictorial and one textual representation and varied their presentation sequence: pictorial followed by textual representation or textual followed by pictorial representation. As the pictorial representation, we used a VR animation displaying a service robot. As the textual representation, we chose an auditory text about robotics. Additionally, we varied whether or not learners received an elaboration prompt before the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *British Journal of Educational Technology* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

learning unit started. We measured learning outcome differentiated for knowledge, comprehension, and application levels. We found beneficial effects for VR animation followed by auditory text on the knowledge level, only. Prompting led to a beneficial effect only on the application level. We found no synergic effect of adequate sequencing and prompting on learning outcome. Hence, depending on the level of learning outcome targeted, different support approaches are advisable. Future research might consider the long-term effects of adequate sequencing and prompting or different approaches to measuring mental models.

KEYWORDS

cognitive learning processes, mental models, prompting, sequencing, virtual reality learning environments

Practitioner notes

What is already known about this topic?

- Learning in Virtual Reality (VR) is a promising approach to present the learning material but does not per se guarantee deep learning and the development of a correct mental model.
- Prior findings imply that presenting pictorial representations first fosters mental model development.
- Beneficial effects of stimulating learners by an elaboration prompt before learning in virtual reality have been reported.
- Previous studies imply that instructional design and activation of learners might have synergetic, beneficial effects on learning outcome.

What this paper adds

- We investigated the effects of sequencing order (VR animation or auditory text first) and prompting (with or without) on different levels of learning outcome.
- We found beneficial effects on the knowledge level when VR animation was presented first, while prompting fostered the application level.
- No synergetic effects of VR animation first and prompting were found as both support approaches seem to impact different levels of learning outcome.

Implications for practice and/or policy

- Depending on the level of learning outcome that should be fostered while learning in VR, different support approaches are advisable.
- Presenting a VR animation first fosters learning of basic concepts and definitions at the knowledge level.
- Using an elaboration prompt before learning in VR enables the learners to process the content deeply.
- Presenting VR animations first can serve as a brief scaffold for the mental models and additional details can be added in the further learning process.

INTRODUCTION

Virtual Reality Learning Environments (VRLEs) offer new possibilities to present multimedia learning content. Previous publications showed a heterogeneous picture of the extent to which VRLEs can be implemented as a deep learning approach (Makransky et al., 2020; Parong & Mayer, 2018; Wu et al., 2020). In our study, we investigated two approaches to stimulate cognitive learning processes and mental model development in VRLEs: Optimization of the presentation sequence of pictorial and textual representations as well as activation of learners before the learning unit by using a prompt.

To understand the potential of VRLEs, they first have to be characterized. A basic distinction when referring to VRLEs is their degree of immersion (Wu et al., 2020). While some publications describe VR as a 3D desktop simulation, others refer to immersive VR that can be entered with the help of head-mounted displays (Wu et al., 2020). Furthermore, VRLEs are categorized based on differing degrees of freedom: allowing simple rotational movements to the possibility to walk around and choose their position in the three-dimensional virtual world. In this study, we refer to an immersive VRLE displayed by a standalone head-mounted display. In this VRLE learners can choose their perspective by moving their head. In the chosen setting learners were not able to interact with the VR animation directly or to move in the virtual world. Most VRLEs use mainly pictorial representations such as pictures or VR animations (e.g. Radianti et al., 2020). Furthermore, textual components can be added. Verbal information might include short narratives, simple labels, or annotations (e.g. Vogt et al., 2021). VRLE might have a high visual load; hence further textual details could be presented as auditory text (e.g. Makransky et al., 2019).

In general, VRLEs are described as motivating and are thought to increase enjoyment while learning (Makransky et al., 2020). *But is learning in VR an effective deep-learning approach?* Literature reviews describe heterogeneous effects of VRLEs on learning outcome (Di Natale et al., 2020; Wu et al., 2020). Some studies describe no benefits or even negative effects of VRLEs (Makransky et al., 2020; Parong & Mayer, 2018, 2020). For instance, Parong and Mayer (2020) found that learning with PowerPoint slides was superior to learning in VR. Ke and Xu (2020) found that learners in the live simulation of chemistry experiments performed better compared to learners in the VR group. In contrast, other studies outline the beneficial effects of learning in VR. For example, Alfadil (2020) found that learners in the VR group outperformed learners in the conventional setting when learning vocabulary. Therefore, presenting learning content in VR does not guarantee a higher learning outcome. This leads to the question: *How can learners be supported to process the learning content in the VRLE?* For instance, Klingenberg et al. (2020) successfully used generative learning strategies to stimulate learning in VR. Furthermore, adding pre-training supports learning in VR (Meyer et al., 2019). Summary prompts lead to higher learning outcome, compared to the control group when learning in VR (Parong & Mayer, 2018). Summarizing these findings, different influencing factors need to be considered to stimulate successful learning in VR.

To understand previous findings, cognitive learning processes need to be considered. The *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2014) describes the active role of the learner as well as information selection, organization, and integration of pictorial and verbal information via two different channels. When the learners have sufficient cognitive resources, they can build a component and a causal mental model and are able to apply the learning content (Mayer et al., 2002). Moreover, the *Integrated Model of Text Picture Comprehension* (IMTPC) describes pictures as analog so they can be directly included into the analog mental model (Schnotz & Bannert, 2003). In contrast, texts consisting of verbal code need to be translated before the integration into the mental model. Additionally, the sequence of presenting textual and pictorial code is crucial for building a coherent mental

model (Eitel & Scheiter, 2015). Presenting general pictures followed by a more specific text including detailed information is more effective than vice versa (Arndt et al., 2019).

Based on the underlying cognitive learning processes, different assumptions can be made about learning in VRLEs. On the one hand, the instructional design has an impact on whether certain cognitive learning processes are fostered (Mayer, 2014). On the other hand, learners' active roles and their strategies have a significant impact on learning outcome (Bannert & Reimann, 2012). To stimulate the processing of the learning content, cognitive prompts have been used successfully in VRLEs (Vogt et al., 2020). Cognitive prompts include different components to stimulate elaboration, and learning strategies to enable the learners to understand and transfer the learning content (Bannert & Reimann, 2012).

The aim of the present study was to investigate whether VR animation or auditory text should be presented first. Additionally, we explore the effect of an elaboration prompt as well as the combination of both factors. To gain further insight into the underlying cognitive learning processes, we used a differentiated measurement of learning outcome (knowledge, comprehension, and application).

THEORETICAL BACKGROUND

Underlying cognitive learning processes in VRLEs

When designing VRLEs the underlying cognitive learning processes need to be considered. Mayer's (2014) CTML describes important assumptions for designing learning material. Two separate channels for information processing exist: textual and pictorial. Using both channels is described as beneficial. As working memory capacity to process learning content is limited, this needs to be considered while designing learning materials or VRLEs (Mayer, 2014). When learners have sufficient resources to process all relevant information they can build a coherent mental model. Based on Mayer et al. (2002), different stages of mental model development can be distinguished. A component model, including definitions or basic concepts, and a causal model, including meaningful connections between the concepts, are built simultaneously. Based on this causal model, learners can deduce consequences from manipulating one component onto another (Mayer et al., 2002). When cognitive overload occurs, no causal model is built. To avoid this overload and to enable learners to process the content deeply including causalities, the instructional design of the VRLE needs to be adapted to these requirements.

This idea of different stages of information processing is also described in Schnotz and Bannert's (2003) IMTPC. Depending on the codality of the learning material, different branches of information processing are described. Both verbal and pictorial representations are first processed sub-semantically and stored as an internal representation based on the given text syntax or picture surface (Schnotz & Bannert, 2003). This allows learners to answer questions at the recall or knowledge level (Bloom, 1956). To be able to understand the meaning of different concepts as well as their connections, learning content needs to be processed semantically. While pictures are analog in nature, they can be directly integrated into the analog mental model after their semantic processing (Schnotz et al., 1999). Verbal information needs to be first translated to be stored in the mental model (Schnotz & Bannert, 2003). Therefore, after semantic processing of verbal information, propositional representations are built. These are then translated and integrated into the mental model. Only when learners process the learning content deeply they can answer questions at higher levels (comprehension and application; Bloom, 1956).

While learning in VRLEs, learners need to process learning content semantically and develop a concept or even a causal mental model. The combination of different

information representations is crucial in fostering these deep-learning processes (Schnotz & Bannert, 2003). As learners' cognitive capacity is limited to processing a certain amount of information simultaneously, presenting all information at once might cause a cognitive overload (Mayer, 2014). *But what is the optimal way to present and combine learning material?* One possibility is to present pictorial and verbal parts of the learning content as a sequence. The chosen presentation format and sequence might impact the development of a coherent mental model substantially.

Foster deep-learning processes through adequate sequencing

To answer the question of how mental model development can be supported by combining pictures and text, Eitel and Scheiter (2015) reviewed $N = 42$ studies. Most of these studies ($n = 19$) outline the beneficial effects of presenting pictorial parts first. When using this sequencing order, pictures can serve as brief scaffolds, and details included in the text can be later added into this scaffold to build a coherent mental model. Recent findings outline the importance of not simply categorizing learning material based on presentation format (pictures or text) but considering how specific the learning content of each representation is. Analyzing different combinations of text or pictures first together with different specificity, they concluded that presenting a brief pictorial overview followed by a more specific text increased learning outcome (Arndt et al., 2019). Furthermore, prior findings outline the benefits of using animations over static pictures when processes or basic mechanisms are also relevant in the performance test (Kühl et al., 2018). Based on previous findings, one can conclude that when learning in a VRLE, VR animations should be presented first followed by a more specific auditory text. Thus, learners are able to build a brief mental model based on the animation and can refine this through adding further details (Arndt et al., 2019). So far, no study has investigated the effect of sequencing pictorial and verbal parts of the learning content on learning outcome using VRLEs.

However, learning is not a passive process and therefore does not only depend on the design and combination of the given learning formats (Mayer, 2014). Whether learners actually engage in integrating the different representations and build comprehensive mental models also depends on the learners. Therefore, the question is how to activate learners effectively.

Activating learners by prompting

One well-established possibility to activate learners, and thereby foster cognitive learning processes, is to use prompts. Using short verbal prompts aims to support the development of a causal mental model (Berthold et al., 2007). Cognitive prompts are characterized as strategy activators as they foster specific cognitive strategies (Bannert & Reimann, 2012). Different realizations of cognitive prompts were described: taking notes, thinking of everyday examples, or generating short summaries (e.g., Bannert & Reimann, 2012; Parong & Mayer, 2018). Using these cognitive prompts aims at fostering deep elaboration of the learning content. Therefore, they are named cognitive or elaboration prompts (Bannert & Reimann, 2012). Additionally, explanatory components might be included in the prompt. These trigger the learners to explain the learning content in their own words, consequently requiring semantic processing of the given content (Berthold et al., 2009).

Prompting has been explored in the context of learning in VR, and different operationalizations investigated in this context. Learners were encouraged to repeat the main concepts of the learning material or were prompted to produce short summaries (Kamarainen et al., 2018; Parong & Mayer, 2018). Overall, the beneficial effects of prompts were described

on simple recall and higher levels of learning outcomes. One possibility to implement prompts is to include them in the instruction. One prior study included an elaboration prompt with an explanatory component before entering the VRLE (Vogt et al., 2020). The beneficial effect of the prompt was limited as learning only increased at the knowledge level. Vogt et al. (2020) attribute this limited effect to missing perceptual or cognitive resources while learning in VR with simultaneously presented pictorial and verbal information. In the present study, we adapted the respective learning materials by sequencing to enable analysis of the interplay between an optimized sequence and prompt. We assume that the sequenced learning material means reduction of perceptual and cognitive load, and therefore learners should have more resources for learning. As learners were relieved by the sequenced material, the expected positive effects of the prompt should become apparent. When the learning sequence starts with animation, a brief scaffold of the mental model can be built, and further details can be added from the text presented after the animation. We assume that this integration process can be fostered by an elaboration prompt, which outlines the connection between the different components and details (Arndt et al., 2019). Therefore, a synergetic effect of adequate sequencing and prompting was expected.

Present study

Based on the theories described and empirical findings, adequate sequencing and prompting as well as their combination are very promising approaches to foster deep learning in VR (Meyer et al., 2019). To date, no study has investigated the beneficial and synergetic effects of adequate sequencing and prompting in VR. This was the aim of the present study. Both support approaches aim to enable learners to process the learning content deeply and answer comprehension and application questions. As answering such questions requires a causal mental model, it can be assumed that a component model is created and the knowledge level is also stimulated (Mayer et al., 2002). Additionally, learners' aptitudes such as their domain-specific prior knowledge or their verbal ability might impact the learning outcome. Therefore, we controlled for these variables in our study.

Our first research question explored the effect of adequate sequencing by presenting either VR animation or auditory text first. As outlined by Eitel and Scheiter (2015), using pictorial representation first, followed by a verbal representation is beneficial for learning outcome. Additionally, presenting a brief pictorial overview followed by a more specific text fostered the development of a coherent mental model (Arndt et al., 2019). Building a brief scaffold based on pictorial information eases the organization and integration of new information into the mental model. Hence, knowledge, as well as comprehension and application, is expected to be promoted. Therefore, we assumed as follows:

H 1 We hypothesized a significant, beneficial effect of presenting VR animation first, followed by a more specific text, compared to vice versa, on all levels of learning outcome.

Additionally, our second research question focused on the effect of prompting on learning outcome. The elaboration prompt triggered deeper elaboration as well as connecting the different concepts. Learners were activated to be able to rephrase the learning content in their own words. To be able to comprehend and apply the learning content, learners need to know basic definitions (Bloom, 1956). Therefore, cognitive processing, particularly the deeper elaboration and transfer of the learning content, was expected to be fostered by the elaboration prompt.

H 2.1 We hypothesized a significant, beneficial effect of the elaboration prompt on all levels of learning outcome, compared to the groups without prompt.

H 2.2 We hypothesized the largest effect of the elaboration prompt on the highest level of learning outcome (application).

Both adequate sequencing and prompting are expected to have beneficial effects on all levels of learning outcome. Prompting might ease the integration of further information into the brief scaffold that was built when presenting VR animation first. Hence, we assume a synergetic effect of adequate sequencing and prompting.

H 3 We hypothesized a significant, beneficial, synergetic effect of both factors: adequate sequence (presenting VR animation first) and prompting compared to the other experimental condition on all levels of learning outcome.

METHOD

A priori power analysis

To conduct our a priori power analysis, we searched for suitable estimations of effect sizes for prompting, sequencing, and their interaction effect in the context of VR learning. As no prior VR study reported suitable effect sizes, we based our a priori power analysis on results in conventional learning environments and thus had only a very brief estimation of the needed sample size. We based our analysis on Eitel and Scheiter (2015; for sequence) and Endres et al. (2017; for prompting). Both studies reported large effect sizes. No study was found that described an appropriate effect size estimate for the interaction between sequencing and prompting. Therefore, we performed a more conservative a priori power analysis by using G*Power version 3.1.9.4 (Faul et al., 2009) to determine the necessary sample size for our planned MANOVA analysis; $f^2(V) = 0.13$ (medium to large effect size based on Cohen, 1988); $\alpha = 0.05$; power $(1 - \beta) = 0.9$. By this, we determined the minimum of required participants with $N = 72$. However, as we were not confident, that the so far reported effect sizes were a very precise estimation of the effect sizes required for our VR study and also no other suitable estimation were available, we extended our sample size up to $N = 81$. By this, we additionally ensured to be able to compensate for a potential drop out due to technical problems.

Sample and study design

In the present study, we tested a sample of participants ($N = 81$), who were mainly psychology students at university. The participants (37% male) were aged between 19 and 38 years ($M_{\text{age}} = 26.08$; $SD_{\text{age}} = 6.71$). We applied a 2×2 between-subject design using an immersive VRLE. The first independent factor was the sequence of presenting learning material: VR animation (pictorial) or auditory text (verbal) first. The second independent factor was whether or not learners received an elaboration prompt. Participants were randomly assigned to one of the four experimental groups: with prompt and auditory text first ($n = 20$), with prompt and VR animation first ($n = 20$), without prompt and auditory text first ($n = 19$), or without prompt and VR animation first ($n = 22$). As dependent variables, we measured learning outcome on the three levels of knowledge, comprehension, and application. For learners' aptitudes, prior knowledge and verbal ability were assessed.

Learning material and support approaches

We extracted our learning unit based on the learning content used in a first-semester university course teaching basics about robotics. The final learning material was critically revised by an expert in robotics to ensure validity and feedback was given by researchers in learning and instruction on its clarity and understandability. The learning material included two different components: auditory text and VR animation (see Appendix). The auditory text contained detailed information about the robot, and the concepts of robotics in general. The VR animation presented the robot and its components (see Figure 1). Learners were able to observe the robot in the VRLE while it conducted its service task. Including a short exploratory phase in the VRLE, the learning unit lasted about 8 min.

The VR environment was programmed in Unity (version 2019.2.0f1). The VRLE included a free-based kitchen environment and the 3D model of REEM robot generated by the manufacturer PalRobotics. The 3D animation of the robot was slightly altered with regard to size and appearance in VR. We used the Oculus Go to display the VRLE. Therefore, learners were able to choose their position in the VRLE but were not able to interact with the robot directly.

To foster processing of the learning content, an elaboration prompt was used. It was successfully used in a prior study to increase the learning outcome (Vogt et al., 2020). The prompt included different parts to stimulate cognitive learning processes: it outlined the connectedness of the presented concepts and contained an explanatory component where learners were asked to prepare a peer explanation in their own words for the learning content (see Appendix).

Questionnaires

Before testing in the laboratory, participants filled in a short *online questionnaire* about their gender, age, educational level, field, prior experiences with VR glasses, and VR applications at home. Additionally, we measured verbal ability by using a subcategory of the IST2000-R (Liepmann et al., 2007). This subcategory was the basic module A (similarities)

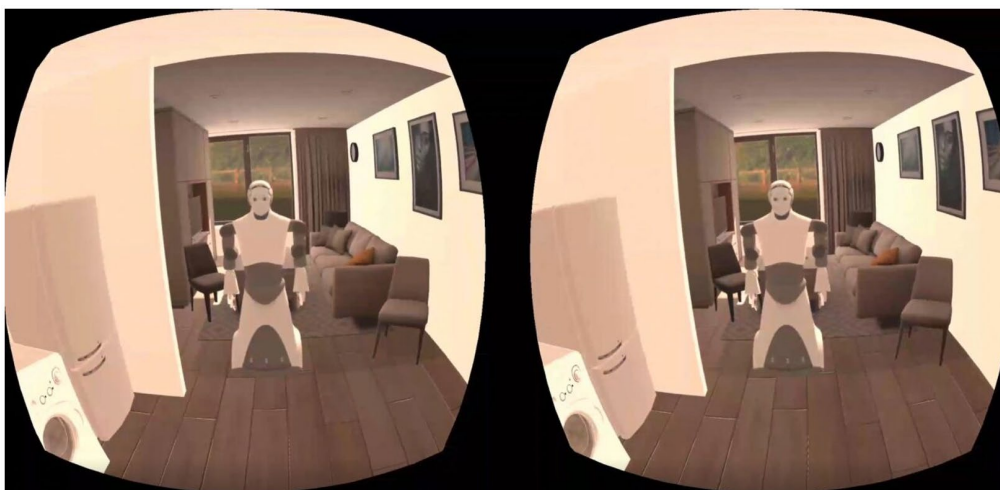


FIGURE 1 Virtual reality learning environment displaying the VR animation of the REEM robot

that consisted of 20 items (item numbers 41–60). Participants were instructed to choose two corresponding words from a list of six words. This decision should rely on an obvious generic term that was compatible with only two words on this list. For example, out of the fictitious list 'fork, butter, necklace, book, biscuit and cigarette' the participants should choose 'butter' and 'biscuit' as these are both food items. Reliability of verbal ability measurement was acceptable with Cronbach's $\alpha = 0.74$ ($CI_{95\%} = 0.62\text{--}0.86$).

We developed our learning questionnaires based on a first-semester lecture in robotics and our questions were additionally critically reviewed by an expert in robotics. Our *pre-questionnaire*, which was answered in the laboratory, included a measurement of prior knowledge to be able to control for differences between the experimental groups. We developed a domain-specific prior knowledge test with seven open questions. For instance, the test included basic questions about relevant technical concepts for robotics (e.g. 'What is the difference between control and regulation in the field of robotics?' or 'Name three different types of sensors that are used in the field of robotics'). To ensure that our scoring was not biased by the rater, we had two independent raters that scored the answers based on a detailed solution scheme. The inter-rater reliability of the two independent raters was high with $r = 0.93$, ($CI_{95\%} = 0.89\text{--}0.96$).

To measure learning outcome, we developed a *post-test* that consisted of twelve open questions. We developed questions for each of the three levels of learning outcome: knowledge, comprehension, application. The four knowledge or recall questions referred to basic definitions (e.g. 'Which term is appropriate to describe that the robot is heading straight towards you?'). To test comprehension level, four questions were included (e.g. 'What can a service robot do in contrast to an industrial robot?') and for application we developed four questions (e.g. 'What happens if the distance between the current pulses in the actuator is shortened?'). Again, the answers were rated by two independent raters who showed a high inter-rater reliability with $r = 0.88$ ($CI_{95\%} = 0.81\text{--}0.93$).

Learners' psychological state of being immersed in the VRLE, also described by the concept of presence (e.g. Slater, 1999), was measured with the respective scale of the Technology Usage Inventory (TUI; Kothgassner et al., 2012). This scale consisted of four items rated on a 7-point Likert scale (e.g. 'Through the virtual simulation I had the feeling of really experiencing the situation'). Reliability of measuring perceived immersion was acceptable, with Cronbach's $\alpha = 0.74$ ($CI_{95\%} = 0.60\text{--}0.88$).

Procedure

The study consisted of two parts. The first was an online questionnaire that was filled in at home. The second part took place in our laboratory at the university as an individual session for each learner. In the laboratory, the participants received a brief introduction about the study by the researcher and signed the informed consent form. The participants were informed that they could withdraw their data or end their participation at any point in the study with no negative consequences. Then the participants filled in the online pre-questionnaire using a laptop provided. Depending on their experimental group, learners received the elaboration prompt. To ensure that they understood the prompt properly, they were asked to repeat the prompt in their own words. Afterwards, the first part of the learning content was presented: either the VR animation or the auditory text. Then, the remaining part of the learning unit was presented. Therefore, all participants viewed the same content; only the order of presentation was altered. Finally, participants filled out the post-questionnaire. Overall, the study lasted about one hour.

Data preparation

To display the online questionnaire, we used the online survey tool Unipark. For data preparation and analyses we used R 3.5.1. For our further analysis of learning test results, we developed a rigorous solution scheme with the help of an expert in robotics. Two independent raters analyzed the answers and scored them based on a pre-defined scoring system for each question. Percentages of correct answers were calculated for both, pre- and post-test, for each level of learning outcome. Additionally, a weighted mean of the three learning outcome scores was calculated to determine the percentages of overall learning outcome for each learner. For verbal ability, the scores were calculated by using the published solution scheme of the IST2000-R and one verbal ability score was deduced for each learner. Furthermore, we checked for outliers and whether assumptions for parametric testing were met (normal distribution, homogeneity of variances and covariances). We based our analysis on the postulated hypotheses. To ensure high power of analysis and to avoid cumulation of alpha error, we used MANOVAs or MANCOVAs instead of multiple ANOVAs or ANCOVAs to determine whether groups differed significantly as recommended by Field (2013). By this, we were able to include more than one dependent variable in each analysis. Additionally, as we used contrasts to analyze our hypotheses. Therefore, we had not to rely on the less powerful post-hoc testing (Field, 2013).

RESULTS

Descriptive results

Learners' domain-specific prior knowledge and verbal ability were on a medium to high level in all four experimental conditions (Table 1). To ensure that no systematic differences between the experimental groups existed, a MANOVA including prior knowledge, verbal ability, and immersion while learning in VR was conducted. Although, we measured presence on a 7-point Likert scale and therefore received ordinal data on perceived immersion, we perform parametric testing with these values as this is a very common approach based on Norman (2010). The four experimental groups did not differ significantly ($F(3,74) = 2.21, p = 0.107$).

Preceding the detailed analyses, based on the descriptive scores of the overall learning outcome, a trend can be described that groups with picture-first performed better than the text-first groups. In detail, the descriptively highest learning outcomes were found for the groups with VR animation first (with or without prompting), followed by auditory text first with prompt (Table 2). The lowest overall learning outcome was achieved in the group with auditory text first and no prompting. No clear pattern can be described when differentiating between all three levels of learning outcome. Descriptively, without controlling for learners' aptitudes, learners with VR animations first achieved higher

TABLE 1 Means and standard deviations of the experimental groups

	VR animation first		Auditory text first	
	<i>M (SD)</i>		<i>M (SD)</i>	
	With prompt	Without prompt	With prompt	Without prompt
	<i>n</i> = 20	<i>n</i> = 22	<i>n</i> = 20	<i>n</i> = 19
Prior knowledge (%)	57.46 (15.41)	58.27 (12.59)	57.91 (13.79)	62.09 (13.42)
Verbal ability (%)	54.07 (14.03)	53.06 (14.54)	61.23 (14.12)	60.09 (9.24)

TABLE 2 Means and standard deviations of the experimental groups on all levels of learning outcome

	VR animation first		Auditory text first	
	<i>M (SD)</i>		<i>M (SD)</i>	
	With prompt	Without prompt	With prompt	Without prompt
	<i>n</i> = 20	<i>n</i> = 22	<i>n</i> = 20	<i>n</i> = 19
Overall (%)	43.71 (11.27)	45.98 (13.14)	42.85 (10.85)	36.92 (13.49)
Knowledge (%)	67.07 (23.64)	68.25 (16.87)	61.23 (14.12)	60.09 (9.24)
Comprehension (%)	39.09 (15.09)	41.25 (25.97)	35.79 (13.05)	34.50 (16.69)
Application (%)	25.00 (12.79)	20.04 (15.55)	28.29 (11.67)	22.50 (11.89)

learning outcomes on knowledge and comprehension levels. Based on the descriptive values, learning outcome on the application level was higher with prompt compared to without prompt.

Impact of sequencing and prompting on learning outcome

Before conducting inferential statistical analysis, we checked for outliers and whether the required assumptions for our planned analysis were met. Based on the Shapiro-Wilk test, the multivariate normal distribution was assumed for relevant variables for each experimental subgroup ($p > 0.060$). The variances were homogenous based on Levene's test ($p > 0.658$). Box's M value of 16.55 was with $p = 0.646$ not significant, hence the observed covariance matrices for the dependent variables were equal across groups.

To test our hypotheses, we performed a MANCOVA ($F(5,74) = 3.27$, $p = 0.010^*$, $\eta^2_{\text{partial}} = 0.194$, Table 3) including learning outcome on all three levels (knowledge, comprehension, and application) with contrasts for presentation sequence and prompting, as well as their interaction. As verbal ability and prior knowledge had a significant impact on learning outcome, we included these as covariates.

In our first hypothesis, we expected a significant impact of sequence (VR animation or auditory text first) on all three learning outcome levels. Based on prior findings, we expected that learning with VR animation first outperforms the group learning with auditory text first. We found a significant, beneficial effect of learning with VR animation first, compared to auditory text first, on the knowledge level only ($H1$; $F(1,68) = 8.89$, $p = 0.004^{**}$, $\eta^2_{\text{partial}} = 0.114$; Table 3). The sequence showed no significant effects on comprehension and application level ($p_{\text{comprehension}} = 0.116$, $p_{\text{application}} = 0.143$). Therefore, learning at the knowledge level was facilitated by presenting VR animation followed by the auditory text, only. Hence, the first hypothesis ($H1$) was only supported partially by the data.

In our second hypothesis, we assumed a beneficial effect of prompting on all levels of learning outcome. We found a significant effect of prompting on application but not for knowledge and comprehension level ($p_{\text{knowledge}} = 0.490$, $p_{\text{comprehension}} = 0.794$, $p_{\text{application}} = 0.018$, $H2.1$). As expected, based on the contrast analysis the elaboration prompt led to a significantly higher learning outcome on the application level and the highest relative effect size ($\eta^2_{\text{partial_application}} = 0.080 > \eta^2_{\text{partial_knowledge}} = 0.007 > \eta^2_{\text{partial_comprehension}} = 0.002$). Therefore, our hypothesis ($H2.2$) was supported by the data.

Our third hypothesis assumed a synergetic effect of both factors: adequate sequencing (VR animation first) and prompting ($H3$). This synergetic effect was not found for the three levels of learning outcome ($F(1,68) = 1.75$, $p > 0.190$, $\eta^2_{\text{partial}} < 0.025$). Hence, based on the present findings, our third hypothesis ($H3$) was rejected.

TABLE 3 Results of the MANCOVA with the different levels of learning outcome and simple contrasts for sequence and prompting, with verbal ability and prior knowledge as covariates

	<i>F</i> (1,68)	<i>p</i>	η^2_{partial}
Knowledge			
Verbal ability	8.37	0.005**	0.110
Prior Knowledge	0.07	0.792	0.001
Prompt	0.45	0.504	0.007
Sequence	8.77	0.004**	0.114
Prompt × Sequence	1.74	0.191	0.025
Comprehension			
Verbal ability	0.01	0.917	<0.001
Prior Knowledge	1.65	0.203	0.024
Prompt	0.12	0.732	0.002
Sequence	2.54	0.116	0.036
Prompt × Sequence	0.58	0.447	0.009
Application			
Verbal ability	0.33	0.568	0.005
Prior Knowledge	5.01	0.029*	0.069
Prompt	5.89	0.018*	0.080
Sequence	2.20	0.143	0.031
Prompt × Sequence	0.40	0.525	0.006

****p* < 0.001; ***p* < 0.01; **p* < 0.05.

DISCUSSION

It is expected that VRLEs will have a substantial impact on education (Meyer et al., 2019; Radianti et al., 2020). However, while some previous publications outline the benefits, others describe no or even negative effects (Makransky et al., 2020; Parong & Mayer, 2018; Wu et al., 2020). Therefore, the question arises about how VRLEs can be implemented as a deep-learning approach. In our study, we used two approaches to stimulate deep learning in VRLEs: We manipulated the presentation sequences (VR animation or auditory text first) and their effect on learning outcome. Additionally, we expected an elaboration prompt to support deep-learning processes when learning with VRLEs. Combining these two approaches (adequate sequencing and prompting) was expected to have a synergetic, beneficial effect on all levels of learning outcome.

Learning outcome depending on presentation sequence of pictorial and verbal information

The first research question addressed the effectiveness of adequate sequencing. Learners in the group with VR animation first outperformed learners in the group with auditory text first on the knowledge level. This beneficial effect of presenting VR animation first is in line with the findings reported in the review of Eitel and Scheiter (2015). However, in the present study, the beneficial effect of presenting VR animation first, was only found for the knowledge level but not the transfer or application of the learning content. Therefore, learners in

the VR animation first group performed better in defining and describing basic concepts but had no advantage when answering comprehension or application questions.

Based on the two-stage theory of mental model construction, one can conclude that learners with VR animation first were supported in building a mental model that can be characterized as a component model (Mayer et al., 2002). When building mental models, two models are developed simultaneously: the component model and the causal model. When the cognitive resources of the learners are insufficient, no causal model is developed. Using a brief pictorial representation (e.g. VR animation) might serve as a scaffold of the main concepts and components. This scaffold seems to support learners to build the component model only.

Having a look at the different learning outcome levels, both groups had a medium to high correct answer rate on the knowledge level, while they scored relatively low for comprehension and application. This further supports the assumption that learners had insufficient resources to process the learning content deeply and to build a causal mental model (Mayer et al., 2002). However, learning successfully does not simply depend on learners' resources or the presentation sequence, but also on their active role while learning. When processing the learning content deeply this active role is a particularly crucial pre-condition for answering questions at the comprehension or application level.

Effects of prompting on different levels of learning outcome

For our second research question, we investigated the activation of learners through an elaboration prompt. Our results imply that the elaboration prompt used in our study supported learners to answer questions at the application level only. No significant beneficial effect was found for knowledge or comprehension. In a prior VR study of Vogt et al. (2020), the same learning material was presented with the VR animation and the auditory text simultaneously, i.e. the learning material did not include a presentation sequence of pictorial and verbal information. In that case, the elaboration prompt fostered learning at the knowledge level only. This was attributed to insufficient resources of learners to build a causal model. Including sequenced material depending on its format enables learners to process the learning content more deeply. The elaboration prompt successfully triggered this deeper elaboration so that learners were able to answer application questions effectively compared to the group without prompt.

One meta-analysis describes how elaboration prompts including an explanatory component have higher impact on transfer or application levels than on the lower levels of learning outcome (Bisra et al., 2018). The authors explain these finding by the fact that, to be able to answer transfer or application questions, learners need to integrate the information from the different representations of the learning content. Answering these questions, therefore, requires a sophisticated mental model to be able to explain the learning content in their own words. Consequently, depending on the questions in the post-test, learners might be more sensitive to this explanatory component compared to answering questions about simple definitions (Bisra et al., 2018). Therefore, the beneficial effects of the elaboration prompt might be found at the application level only.

Synergetic effect of adequate sequence and prompting on learning outcome

Our third research question explored a synergetic effect of both presenting VR animation first and prompting on learning outcome. This assumption was not supported by the present

results. We expected both supportive means to foster learning outcome at all levels as the beneficial effects of prompting are highly dependent on prior knowledge. Depending on the presentation sequence of auditory text and VR animation, we assumed a brief scaffold of the learning content was built and this was more or less efficient. This scaffold can be described as the pre-existing mental model and can therefore be defined as prior knowledge. Hence, supportive means were assumed to show synergetic effects (Leppink et al., 2012). However, presenting VR animation first supported learners at the knowledge level, while prompting supported learners at the application level. Based on these findings, no synergetic effect was found as presentation sequence and prompting supported very different cognitive sub-processes of learning. To promote the knowledge level, superficial processing at the syntax level of learning material is sufficient (Bloom, 1956; Schnotz & Bannert, 2003). The brief pictorial overview through presenting the VR animation fostered this level of learning outcome. In the IMTPC this reflects the sub-semantic processing of the given learning content (Schnotz & Bannert, 2003). Building an internal (but not semantic) representation of the content was fostered by presenting VR animation first. By contrast, the elaboration prompt fostered semantic processing of the learning outcome. Based on the IMTPC, integration of information into the mental model by processes of model inspection was triggered by the elaboration prompt during the learning unit (Berthold et al., 2009). As the supported cognitive sub-processes are substantially different and support learners on the sub-semantic or the semantic level, the missing synergetic effect is in line with the results from the first two research questions in this study.

Limitations and implications for future research

In the present study, we split up the learning unit depending on its format (auditory text and VR animation). However, one could alter presentation sequence, formats, and specificity in further combinations. Besides presenting verbal information as auditory text, the written text could be used. For example, verbal annotations could be added to the VR animation to support learners to process information (Vogt et al., 2020).

We planned our study and the necessary sample size based on previous findings of Eitel and Scheiter (2015; for sequence) and Endres et al. (2017; for prompting). However, we uncovered smaller effect sizes than initially expected. Therefore, our non-significant results of prompting and sequencing on some level of learning outcome might be also due to a lack of statistical power (Faul et al., 2009). The effect sizes reported in our study could now be used to calculate a priori power analysis for future VR studies.

One limitation of this study was the short duration of the learning unit. Having a look at previous instructional design research (see Mayer, 2014), very short learning units are rather common to investigate different instructional design research questions and their effects on different levels of learning outcome (e.g. Mayer & Anderson, 1991). More recent VR studies also used rather short learning units of 7 to 12 min (e.g. Albus et al., 2021; Parong & Mayer, 2018; Vogt et al., 2021). We nevertheless would recommend that future studies might consider investigating longer learning units to investigate even more complex learning processes.

Furthermore, there was only very little delay between the learning session and the performance test. Hence, learners might lack time to process the new information deeply and to integrate new information into their mental model. This was in line with previous publications (see Albus et al., 2021; Mayer, 2014) which asked learners to answer questions immediately after learning and uncovered effects on different levels of learning outcome. Consequently, missing effects on learning outcome cannot be attributed to the study design and the short time between the learning session and the performance test. However, in future studies,

the long-term effects of sequencing and prompting on higher learning outcomes might be analyzed.

Using prompts before entering the VRLEs showed beneficial effects. Further guidance through prompts either before learning in VR or while learning in VR could be implemented (e.g. Johnson et al., 2013). With altered prompts, learners could be supported on strategic levels, for instance by using metacognitive prompts. This could extend the beneficial effect of the prompts used to learning outcome at all levels.

In the given VRLE, learners were able to choose their perspective and to observe the VR animation, but they were not able to walk around or interact in VR. In future studies, additional interactive elements, further degrees of freedom and possibilities to explore the learning content in a more self-paced and active way are might be investigated (Wu et al., 2020). In addition, we used a highly controlled setting, as only one learner was present in the lab during our study. Future studies could investigate group settings of VR learning as well as possible interaction between learners or teachers and learners.

We used a differentiated measurement of learning outcome. The questions developed for comprehension and application levels were rather challenging and should be adapted in future studies. We also found evidence that calls Bloom's (1956) hierarchy of learning into question. Therefore, different measurements of learning processes and outcome are recommended. For instance, process data should be collected in future studies to investigate learning or exploration strategies, particularly in more self-paced and interactive VRLEs.

In this study, as well as in the previous study with the same learning material, the challenge of learners' limited cognitive resources and capacity has been mentioned. Therefore, it is advisable to explore the cognitive load while learning. In particular, a differentiated measurement including items for germane, intrinsic, and extraneous load could deliver further insights into learners' cognitive load, as well as their invested resources while learning. As the active role of the learners is crucial, motivational concepts should also be considered and explored in future work.

CONCLUSION

The present study aimed to contribute to the question of how to turn learning in VR into a deep-learning approach. Based on the findings, depending on the level of learning outcome targeted, different support approaches are advisable. For learning basic concepts or definitions, presenting a VR animation followed by a more specific auditory text is a promising approach. To foster deeper processing of the learning content, an elaboration prompt before entering the VRLE is advisable, as it supports learners to build connections between the concepts and to process the content semantically. In addition to the present findings, other influencing factors need to be investigated to develop adequate instructional designs for VRLEs. To gain a deeper understanding of cognitive learning processes, further underlying processes should be measured and explored. Learners' aptitudes should also be taken into account. Previous findings should be replicated and further explored in other disciplines and contexts to reach the full potential of VRLEs.

ACKNOWLEDGMENTS

We would like to thank our students and the experts consulted in this study.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ETHICS STATEMENT

The study was carried out in accordance with the Declaration of Helsinki. Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. The data were used pseudonymously, and participants were aware that they had the chance to withdraw their data at any point in the study.

DATA AVAILABILITY STATEMENT

The relevant data from the current dataset can be made available by contacting the corresponding author.

REFERENCES

- Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, 166, 104154. <https://doi.org/10.1016/j.compedu.2021.104154>
- Alfadi, M. (2020). Effectiveness of virtual reality game in foreign language vocabulary acquisition. *Computers & Education*, 153, 103893. <https://doi.org/10.1016/j.compedu.2020.103893>
- Arndt, J., Schüler, A., & Scheiter, K. (2019). Investigating the influence of simultaneous—versus sequential—text-picture presentation on text-picture integration. *The Journal of Experimental Education*, 87(1), 116–127. <https://doi.org/10.1080/00220973.2017.1363690>
- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. <https://doi.org/10.1007/s11251-011-9167-4>
- Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, 37(4), 345–363. <https://doi.org/10.1007/s11251-008-9051-z>
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17(5), 564–577. <https://doi.org/10.1016/j.learninstruc.2007.09.007>
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, 30(3), 703–725. <https://doi.org/10.1007/s10648-018-9434-x>
- Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook I: The cognitive domain*. David McKay Co Inc.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Earlham 628 Associates.
- Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, 51(6), 2006–2033. <https://doi.org/10.1111/bjet.13030>
- Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, 27(1), 153–180. <https://doi.org/10.1007/s10648-014-9264-4>
- Endres, T., Carpenter, S., Matrin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13–20. <https://doi.org/10.1016/j.learninstruc.2016.11.010>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage Publication.
- Johnson, A. M., Butcher, K. R., Ozogul, G., & Reisslein, M. (2013). Learning from abstract and contextualized representations: The effect of verbal guidance. *Computers in Human Behavior*, 29(6), 2239–2247. <https://doi.org/10.1016/j.chb.2013.05.002>
- Kamarainen, A. M., Thompson, M., Metcalf, S. J., Grotzer, T. A., Tutwiler, M. S., & Dede, C. (2018). Prompting connections between content and context: blending immersive virtual environments and augmented reality for environmental science learning. In D. Beck, C. Allison, & L. Morgado (Eds.), *Communications in Computer and Information Science. Immersive Learning Research Network: 4th International Conference (ILRN 2018), Missoula, MT, USA, June 24-29, 2018, proceedings* (Vol. 840, pp. 36–54). Springer. https://doi.org/10.1007/978-3-319-93596-6_3
- Ke, F., & Xu, X. (2020). Virtual reality simulation-based learning of teaching with alternative perspectives taking. *British Journal of Educational Technology*, 51(6), 2544–2557. <https://doi.org/10.1111/bjet.12936>
- Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology*, 51(6), 2115–2138. <https://doi.org/10.1111/bjet.13029>

- Kothgassner, O. D., Felnhöfer, A., Hauk, N., Kastenhofer, E., Gomm, J., & Kryspin-Exner, I. (2012). *TUI: Technology usage inventory*. FFG. https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf
- Kühl, T., Navrátil, S. D., & Münzer, S. (2018). Animations and static pictures: The influence of prompting and time of testing. *Learning and Instruction*, 58, 201–209. <https://doi.org/10.1016/j.learninstruc.2018.07.006>
- Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P. M., & Berger, M. P. F. (2012). Self-explanation in the domain of statistics: An expertise reversal effect. *Higher Education*, 63(6), 771–785. <https://doi.org/10.1007/s10734-011-9476-1>
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R*. Hogrefe.
- Makransky, G., Andreassen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*. Advance Online Publication. <https://doi.org/10.1037/edu0000473>
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Mayer, R. (2014). *The Cambridge handbook of multimedia learning*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484. <https://doi.org/10.1037/0022-0663.83.4.484>
- Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, 8(3), 147–154. <https://doi.org/10.1037/1076-898X.8.3.147>
- Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140, 103603. <https://doi.org/10.1016/j.compedu.2019.103603>
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797. <https://doi.org/10.1037/edu0000241>
- Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, 37(1), 226–241. <https://doi.org/10.1111/jcal.12482>
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schnotz, W., Böckeler, J., & Gröndziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, 14(2), 245–265. <https://doi.org/10.1007/BF03172968>
- Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5), 560–565. <https://doi.org/10.1162/105474699566477>
- Vogt, A., Albus, P., & Seufert, T. (2021). Learning in virtual reality: Bridging the motivation gap by adding annotations. *Frontiers in Psychology*, 12, 881. <https://doi.org/10.3389/fpsyg.2021.645032>
- Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2020, August). *Learning about Robotics in Virtual Reality*. Paper presented at the meeting of EARLI - SIG 6 & 7 Conference, Tübingen/online.
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005. <https://doi.org/10.1111/bjet.13023>

How to cite this article: Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome. *British Journal of Educational Technology*, 00, 1–19. <https://doi.org/10.1111/bjet.13104>

APPENDIX

In the following, we provide further information about the learning material and the most important concepts included in the narration and the virtual reality animation. Furthermore, we provide more details about the different levels of learning outcome and the elaboration prompt.

Learning unit: Auditory text

Two main topics were included in the auditory learning unit: First, it included details about the robot, such as its technical specification; second, it included more abstract information for instance the definition of kinematic. To encourage learners' engagement with the topic, the robot was the narrator (*Hello, I am a service robot for home use.*).

To begin, the robot described its size, weight, and transport platform at the back. It also explained the difference between service and industrial robots. The robot explained its main function: cleaning floors. It also described the necessary processes and the related technical components it needs for its cleaning task. In addition, the robot outlined its functionality of certain sensors and the existing safety protocols (*First, I capture my environment through various sensors and a camera...*).

Part of this explanation related to further details about data processing, its fusion, and its translation into corresponding movements and actions of the robot. The robot described its possible motion trajectory and went into more detail about the link to concepts of kinematics (*I can move in a straight line, which is also described by the term...*).

The robot explained further details about the motion system and the underlying control circuit. One topic, for example, was a special controller (*PID controller*), which influenced the movement speed of the robot. Here, the functionality of the controller and the technical interaction with the motion-executing components were described. Based on this information, learners had the opportunity to learn about various processes related to robotics and to develop a deeper understanding of the interrelationships of individual components. Another abstract topic was introduced at the end of the learning unit: the difference between local and global navigation. It was also explained what consequences the respective type of navigation would have for the cleaning task (*...in global navigation, I have a more comprehensive environment model and try to reach my cleaning goal without driving into unnecessary dead ends...*).

Learning unit: Virtual reality animation

The virtual reality learning environment included an animated 3D model of the humanoid REEM robot of the manufacturer PalRobotics. We conducted the study in cooperation with a further study, investigating human–robot interaction. Hence, the robot model was adapted to be slightly larger than its real counterpart (1.75 m instead of 1.6 m). Learners were able to see the different components of the robot and its moving trajectory. Additionally, they were observing the robot in action while executing its service task. We programmed the virtual reality learning environment with Unity (version 2019.2.0f1). To display the virtual reality we used the standalone head-mounted display Oculus Go. The learners were able to choose their exploration perspective and move in the virtual reality. Hence, learners could peek around the corner to watch the robot while performing its service task in the virtual kitchen. They were able to look up and down to focus on different parts of the robot for instance its face or its chassis. Additionally, learners were able to yaw to be able to focus the robot all

the time while it was moving around. However, the learners were not able to manipulate the robot.

Learning unit: Different levels of learning outcome

The learning unit provided information to reflect the different learning outcomes based on the taxonomy of Bloom (1956). The robot's narration as well as the virtual reality animation of the robot contained information such as facts about the height and weight of the robot. This reflected the learning outcome on the knowledge level (Bloom, 1956). Furthermore, based on the provided details, learners were able to compare different concepts and to name their distinction (e.g. global and local navigation). A deeper understanding was stimulated by providing facts and processes in both, the narration and the virtual reality animation to enable the learners to transfer this knowledge to new problems (e.g. *What happens if the distance between the current pulses in the drive is shortened?*).

Elaboration prompt

Depending on their experimental conditions, the investigator read out the elaboration prompt before the learning unit started. To ensure, that learners understood the elaboration prompt, they were kindly asked to repeat it out loud in their own words to the investigator. When necessary the prompt was explained or repeated again. The elaboration prompt was:

You will now get to know the robot better. You will also learn some details about robotics. Pay particular attention to concepts and terms that are new to you and how they are related to each other. Imagine that you have to explain them to someone else who has no prior knowledge in this field.

Using this elaboration prompt aimed at fostering deep elaboration of the learning content by outlining potential connections between the different concepts. Hence, it stimulated cognitive learning processes, and activated learners to elaborate the learning content deeply (Bannert & Reimann, 2012). Based on prior findings, prompts are more effective when these activate learners to repeat the learning content in their own words (Berthold et al., 2009). Hence, the included explanatory components aimed to trigger the learners to prepare a peer explanation. To do so, learners require to process the learning content deeply and thus semantically (Berthold et al., 2009).

Appendix A

In the following, we provide further information about the learning material and the most important concepts included in the narration and the virtual reality animation. Furthermore, we provide more details about the different levels of learning outcome and the elaboration prompt.

Learning unit: Auditory text

Two main topics were included in the auditory learning unit: First, it included details about the robot, such as its technical specification; second, it included more abstract information for instance the definition of kinematic. To encourage learners' engagement with the topic, the robot was the narrator (*'Hello, I am a service robot for home use.'*).

To begin, the robot described its size, weight, and transport platform at the back. It also explained the difference between service and industrial robots. The robot explained its main function: cleaning floors. It also described the necessary processes and the related technical components it needs for its cleaning task. In addition, the robot outlined its functionality of certain sensors and the existing safety protocols (*'First, I capture my environment through various sensors and a camera...'*).

Part of this explanation related to further details about data processing, its fusion, and its translation into corresponding movements and actions of the robot. The robot described its possible motion trajectory and went into more detail about the link to concepts of kinematics (*'I can move in a straight line, which is also described by the term...'*).

The robot explained further details about the motion system and the underlying control circuit. One topic, for example, was a special controller (*PID controller*), which influenced the movement speed of the robot. Here, the functionality of the controller and the technical interaction with the motion-executing components were described. Based on this information, learners had the opportunity to learn about various processes related to robotics and to develop a deeper understanding of the interrelationships of individual components. Another abstract topic was introduced at the end of the learning unit: the difference between local and global navigation. It was also explained what consequences the respective type of navigation would have for the cleaning task (*'...in global navigation, I have a more comprehensive environment model and try to reach my cleaning goal without driving into unnecessary dead ends...'*).

Learning unit: Virtual reality animation

The virtual reality learning environment included an animated 3D model of the humanoid REEM robot of the manufacturer PalRobotics. We conducted the study in cooperation with a further study, investigating human-robot interaction. Hence, the robot model was adapted to be slightly larger than its real counterpart (1.75 m instead of 1.6 m). Learners were able to see the different components of the robot and its moving trajectory. Additionally, they were observing the robot in action while

executing its service task. We programmed the virtual reality learning environment with Unity (version 2019.2.0f1). To display the virtual reality we used the standalone head-mounted display Oculus Go. The learners were able to choose their exploration perspective and move in the virtual reality. Hence, learners could peek around the corner to watch the robot while performing its service task in the virtual kitchen. They were able to look up and down to focus on different parts of the robot for instance its face or its chassis. Additionally, learners were able to yaw to be able to focus the robot all the time while it was moving around. However, the learners were not able to manipulate the robot.

Learning unit: Different levels of learning outcome

The learning unit provided information to reflect the different learning outcomes based on the taxonomy of Bloom (1956). The robot's narration as well as the virtual reality animation of the robot contained information such as facts about the height and weight of the robot. This reflected the learning outcome on the knowledge level (Bloom, 1956). Furthermore, based on the provided details, learners were able to compare different concepts and to name their distinction (e.g. global and local navigation). A deeper understanding was stimulated by providing facts and processes in both, the narration and the virtual reality animation to enable the learners to transfer this knowledge to new problems (e.g. *What happens if the distance between the current pulses in the drive is shortened?*).

Elaboration prompt

Depending on their experimental conditions, the investigator read out the elaboration prompt before the learning unit started. To ensure, that learners understood the elaboration prompt, they were kindly asked to repeat it out loud in their own words to the investigator. When necessary the prompt was explained or repeated again. The elaboration prompt was: *'You will now get to know the robot better. You will also learn some details about robotics. Pay particular attention to concepts and terms that are new to you and how they are related to each other. Imagine that you have to explain them to someone else who has no prior knowledge in this field.'*

Using this elaboration prompt aimed at fostering deep elaboration of the learning content by outlining potential connections between the different concepts. Hence, it stimulated cognitive learning processes, and activated learners to elaborate the learning content deeply (Bannert & Reimann, 2012). Based on prior findings, prompts are more effective when these activate learners to repeat the learning content in their own words (Berthold, Eysink, & Renkl, 2009). Hence, the included explanatory components aimed to trigger the learners to prepare a peer explanation. To do so, learners require to process the learning content deeply and thus semantically (Berthold, Eysink, & Renkl, 2009).

Appendix B - Additional analysis

Additional analysis

All studies conducted during this thesis were part of projects that focused on additional research questions. Therefore, only some findings were published in the included journal papers. As some further findings are related to the rational of this thesis, these are provided in more detail in this section.

As part of the related research projects further potential influencing learner's characteristics were investigated. As already outlined in this thesis, different learners characteristics might impact learning outcome. To broaden the insights into the importance of learners characteristics and potential interactions with the given support approaches more details are provided.

Study 1 - Learning in Virtual Reality: Bridging the Motivation Gap by Adding Annotations

The publication focussed on the analysis directly related to the postulated hypothesis (Vogt, Albus, & Seufert, 2021). To deduce further details about the underlying cognitive learning processes that were affected by the annotations additional explorative analysis were conducted and the main findings of these analysis are reported.

Briding the motivation gap

In the publication, the moderating effect of intrinsic motivation on overall learning outcome was focussed. To gain further insight into this effect, each level of learning outcome has been explored. Compared to the knowledge and comprehension level, the application level did not show this moderating effect (see Figure 1 and table 1). The strongest moderating effect was found on comprehension level (see Figure 2 and table 2).

Table 1.: Linear model with moderating effect on application level

Variable	Coefficient	SE	t-value	p
(Intercept)	-0.10	0.16	-0.65	0.519
Group	0.23	0.23	0.98	0.33
Prior knowledge	0.13	0.04	3.23	0.002
Intrinsic motivation	0.16	0.10	1.57	0.122
Group * intrinsic motivation	-0.05	0.15	-0.33	0.751

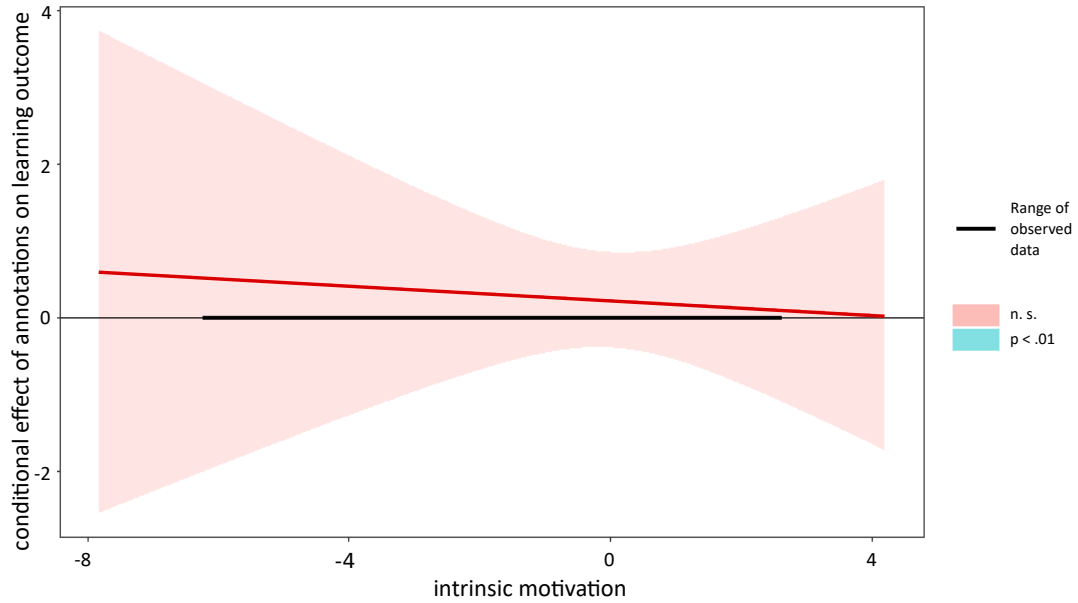


Figure 1.: Johnson-Neyman interval displaying the missing moderating effect of intrinsic motivation on application level

Table 2.: Linear model with moderating effect on comprehension level

Variable	Coefficient	SE	t-value	p
(Intercept)	-0.08	0.15	-0.53	0.597
Group	0.21	0.21	0.97	0.338
Prior knowledge	0.13	0.04	3.32	0.001
Intrinsic motivation	0.29	0.09	3.06	0.003
Group * intrinsic motivation	-0.40	0.14	-2.92	0.005

Learner's characteristics

Since working memory capacity is limited and in complex learning environments a lot of information has to be processed simultaneously, it is plausible that learning outcome also depends to a substantial degree on the working memory capacity of the learner (Baddeley, 1992). Furthermore, as one main feature of VRLEs is displaying the spatial configuration of the learning content in an immersive way, and the impact of spatial ability on learning outcome can be expected in the context of learning in VR. In line with this idea, working memory capacity and spatial ability were examined in this project related to Study 1. In this study $N = 107$ learners were learning in VRLEs providing information about seawater desalination by using VR animations, textual labels, and narration (for further details and findings of this study see Albus et al. (2021)). A positive relation between learning outcome and working memory capacity was expected as well as a beneficial or moderating effect of spatial ability. However, based on the reported findings in this study the importance of working memory capacity and spatial ability was not supported empirically (Albus et al., 2021). To gain further insight into the importance of these factors, future studies should again systematically examine the

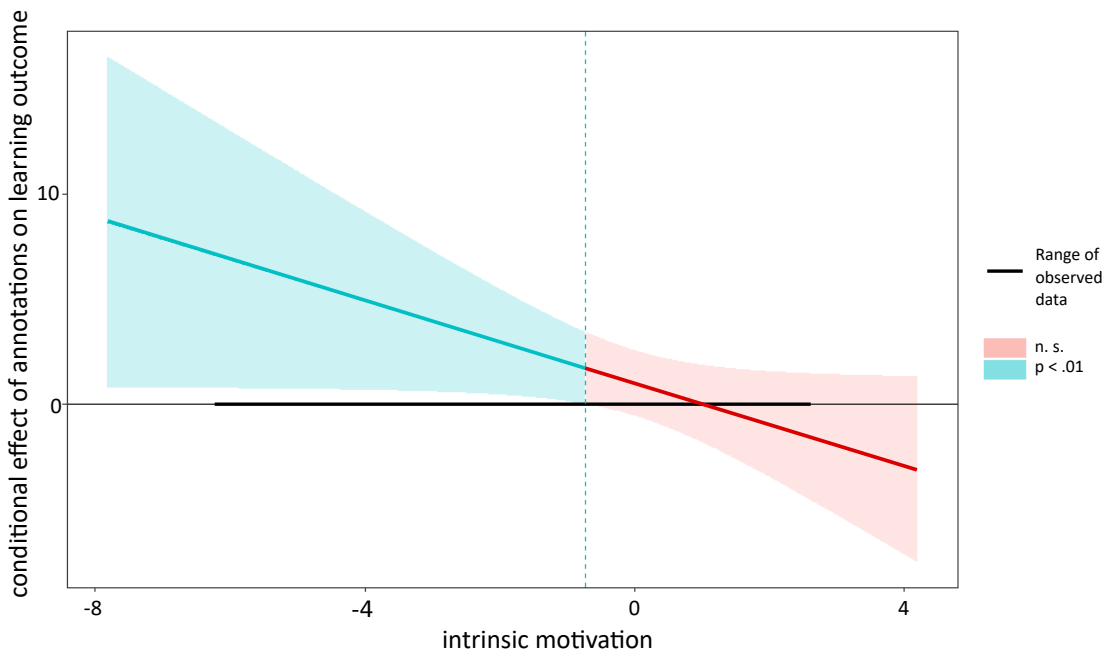


Figure 2.: Johnson-Neyman interval displaying the moderating effect of intrinsic motivation on comprehension level

extent to which the potential influence of these factors also depends on the learning material or instructional design of the VRLE.

Additional Study 1 - Supporting learners by color-coding

Short summary

One further external support approach for learning in VR was investigated (Vogt, Albus, Montag, et al., 2021). In this study, a non-immersive virtual reality learning environment was used. Previous findings on learning in VR imply beneficial effects of including visual cues or highlights in the VRLE (Vogt, Albus, & Seufert, 2021; Albus, Vogt, & Seufert, 2021). When different representations are present in the virtual reality learning environment, learning can be supported if corresponding learning content is appropriately highlighted in textual and pictorial representations. One approach to facilitate the mapping of corresponding elements is color-coding (Mayer, 2014). Here, corresponding elements are marked as belonging to each other by the same color scheme. This was intended to support different cognitive learning processes. On the one hand, color-coding supports finding the corresponding elements and therefore eases visual search similar to annotations (Ozcelik et al., 2010). In addition, like annotations, it supports the integration of different information in the subsequent processing procedures (Mayer, 2014).

In this VR study with $N = 54$ learners, the learning content on the topic of statics was displayed by different presentations formats in the VRLE: Learners were able to learn about different basic formulas related to statics on whiteboards and explore these relationships

between different parameters directly and interactively in the VRLE. For this purpose, learners could place different weights on beams that differed in material, diameter, or length. Each of the four chapters included a short interactive quiz. The central learning contents of the different chapters were then presented again on an overview board. The experimental group was additionally supported by color-coding to map corresponding elements on the different panels and in the graphics, as well as virtual reality animations. The control group, on the other hand, had no such support for the assignment of the elements. The effects of color-coding were investigated on learning outcomes. The learning outcome was measured in a differentiated way by using the levels of Bloom (1956) knowledge, comprehension, and applications. Different learner characteristics were examined. For instance, domain-specific prior knowledge was accessed. A beneficial effect of color-coding compared to the control group was expected on the overall learning outcome (H1). Based on theoretical considerations, particularly learning on the application level was assumed to be increased (H2).

Main findings

Before conducting inferential statistics to test the postulated hypothesis, potential covariates were considered.

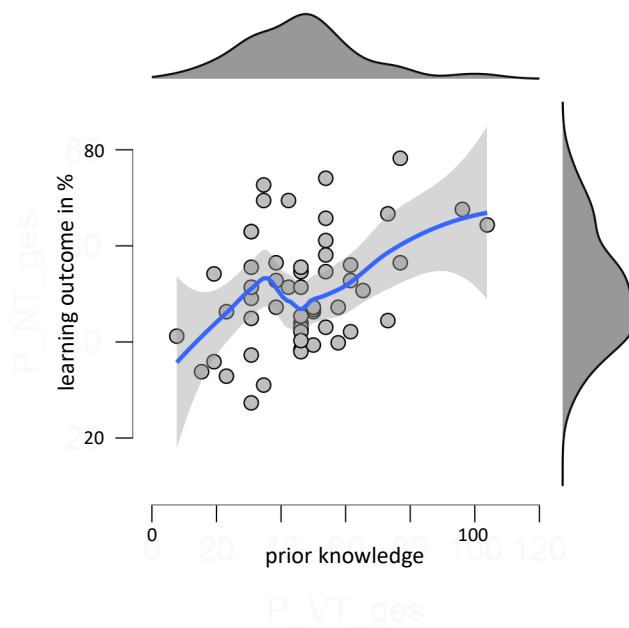


Figure 3.: Relationship of prior knowledge and overall learning outcome

A positive relationship between learning outcome was found for prior knowledge (see Figure 3). Thus, prior knowledge was included into the subsequent analysis for testing the hypothesis.

Analysing the effect of color-coding on the overall learning outcome, no significant effect of color-coding was found (H1; see table 3). Considering the covariate prior knowledge, a positive effect of color coding on the application level was found (H2), see table 4.

Table 3.: Results of the ANCOVA including experimental group and the covariate prior knowledge on overall learning outcome

Cases	Sum of Squares	df	Mean Square	F	p	η^2
group	78.587	1	78.587	0.667	0.418	0.011
P_VT_ges	1202.404	1	1202.404	10.205	0.002	0.162
Residuals	6126.974	52	117.826			

Table 4.: Results of the ANCOVA including experimental group and the covariate prior knowledge on learning outcome on the application level

Cases	Sum of Squares	df	Mean Square	F	p	η^2
group	1502.652	1	1502.652	4.819	0.033	0.077
P_VT_ges	1785.778	1	1785.778	5.727	0.020	0.092
Residuals	16215.092	52	311.829			

Appendix C - Exemplary R code

Data analysis with R

Exemplary R code for data preparation and statistical analysis

This is exemplary code has been implemented using R-4.1.0, RStudio 1.1.463 and R Markdown.

```
library(ggplot2)
library(ggvis)
library(corrplot)
library(ggcorrplot)
library(psych)
library(moments)
library(DescTools)
library(lawstat)
library(pander)
library(jtools)
library(MVN)
library(interactions)
library(wesanderson)
library(heplots)
library(rstatix)
library(ggpubr)
```

Generating random data frames

```
dat <- data.frame(replicate(10, sample(1:100, 100, rep=TRUE)))
```

Naming columns

```
names(dat) <- c('class', 'age', 'IQ', 'prior_knowledge', 'motivation', 'ability', 'learning_outcome', 'knowledge', 'comprehension', 'application')
```

Add experimental groups and case numbers

```
a <- sample(1:4, 100, rep=TRUE)
dat$group <- a
b <- seq(1, 100, by = 1)
dat$case <- b
c <- sample(0:1, 100, rep=TRUE)
dat$group.dummy <- c
```

Brief overview

```
head(dat)
```

```
##      class age IQ prior_knowledge motivation ability learning_outcome knowledge
## 1      81  37 45                76          99      59                70        3
## 2      28  25 46                85          86      15                44       72
## 3      42  25 14                55           2       2                64        8
## 4      85  88 83                86          56      76                72       12
## 5      46  28 23                59          22      77                51       11
## 6      99  41 35                52          46      24                51       37
##      comprehension application group case group.dummy
## 1                12            14      2      1          1
## 2                94            93      4      2          1
## 3                19            43      3      3          0
## 4                 8            91      4      4          1
## 5                66            60      3      5          1
## 6                29            77      4      6          1
```

Check matching cases and merging of data frames

```
match <- function(h, i) {
  return(list(h[!(h %in% i)], i[!(i %in% h)]))
}

dat_2 <- data.frame(replicate(4,sample(1:100,100,rep=TRUE)))
names(dat_2) <- c('case', 'attention', 'time', 'interaction')
dat_2$case<- b
head(dat_2)
```

```
##      case attention time interaction
## 1      1         6   75           9
## 2      2        23   74          34
## 3      3        98   27          37
## 4      4        28   25          28
## 5      5        30   82          99
## 6      6        93   12          11
```

```
d_ges <- merge(dat,dat_2,by="case")
head(d_ges)
```

```
##      case class age IQ prior_knowledge motivation ability learning_outcome
## 1      1      81 37 45              76          99      59              70
## 2      2      28 25 46              85          86      15              44
## 3      3      42 25 14              55          2       2              64
## 4      4      85 88 83              86          56      76              72
## 5      5      46 28 23              59          22      77              51
## 6      6      99 41 35              52          46      24              51
##      knowledge comprehension application group group.dummy attention time
## 1              3              12              14      2              1          6      75
## 2              72              94              93      4              1         23      74
## 3              8              19              43      3              0         98      27
## 4              12              8              91      4              1         28      25
## 5              11              66              60      3              1         30      82
## 6              37              29              77      4              1         93      12
##      interaction
## 1              9
## 2              34
## 3              37
## 4              28
## 5              99
## 6              11
```

Check data types

```
typeof(d_ges$group)
```

```
## [1] "integer"
```

```
d_ges$group <- as.factor(d_ges$group)
d_ges$group.dummy <- as.factor(d_ges$group.dummy)
```

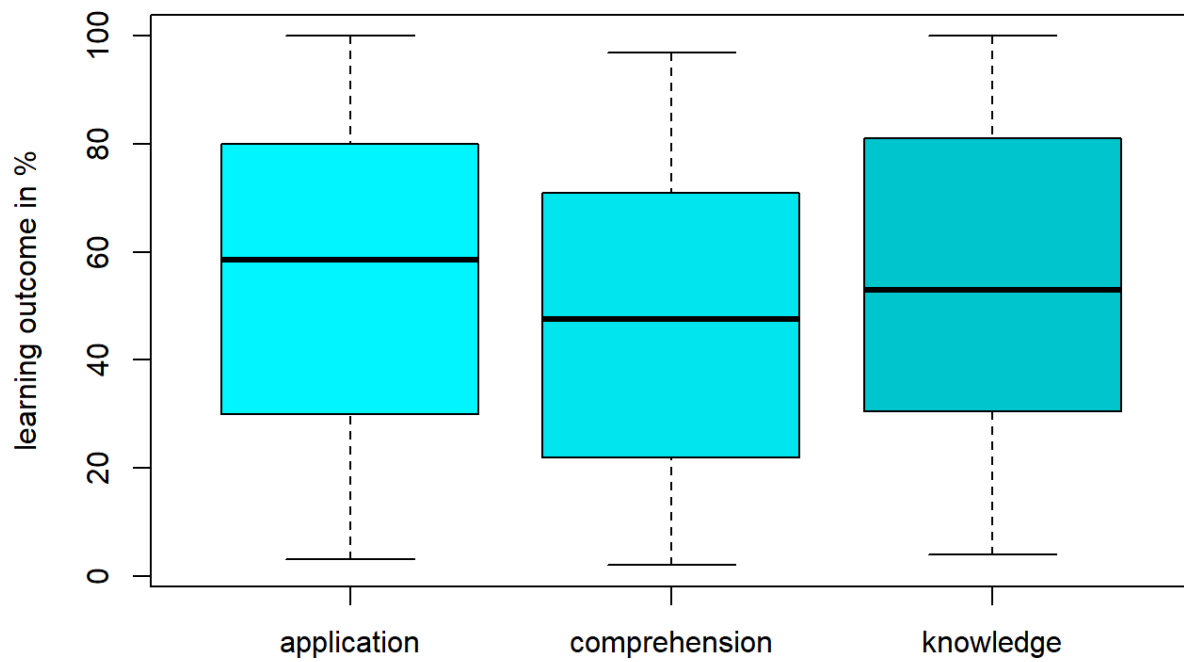
Exploring the dataset

```
summary(d_ges)
```

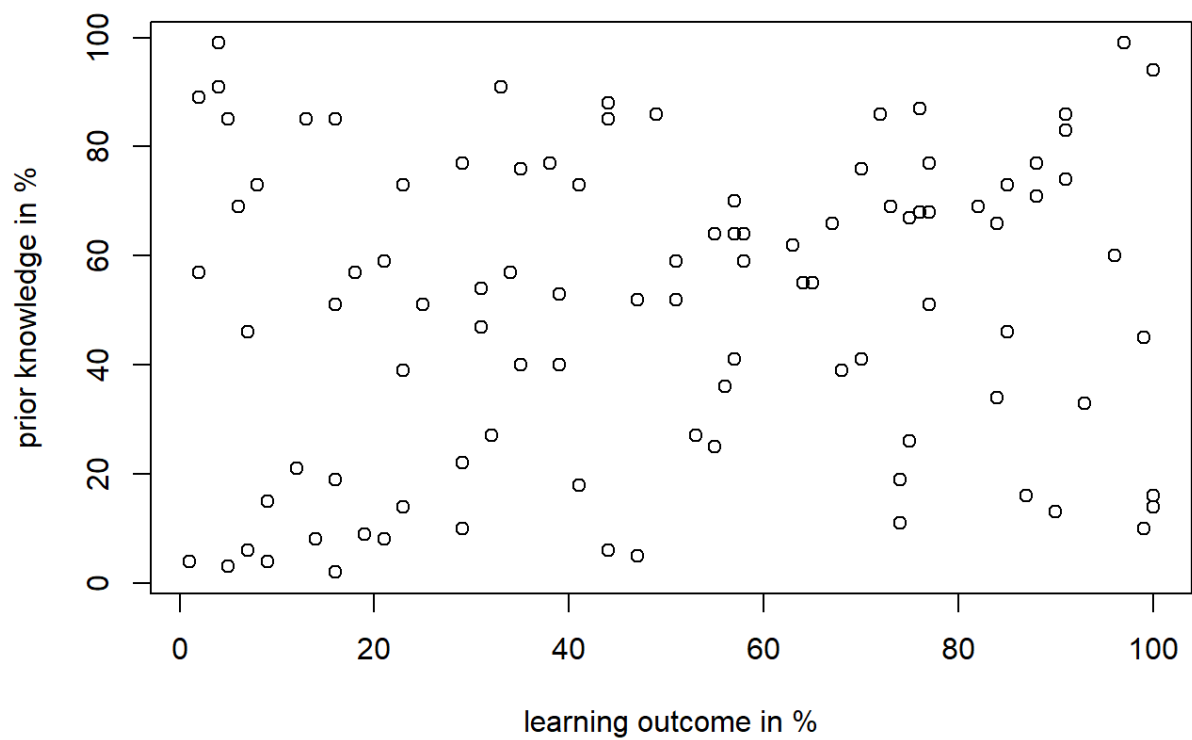


```
##      case      class      age      IQ
##  Min.   : 1.00   Min.   : 1.00   Min.   : 2.00   Min.   : 1.00
## 1st Qu.: 25.75   1st Qu.: 26.75   1st Qu.:27.75   1st Qu.: 25.50
## Median : 50.50   Median : 59.00   Median :50.00   Median : 53.00
## Mean   : 50.50   Mean   : 53.66   Mean   :49.67   Mean   : 52.28
## 3rd Qu.: 75.25   3rd Qu.: 78.50   3rd Qu.:73.00   3rd Qu.: 79.00
## Max.   :100.00   Max.   :100.00   Max.   :98.00   Max.   :100.00
## prior_knowledge motivation ability learning_outcome
##  Min.   : 2.00   Min.   : 1.00   Min.   : 1.00   Min.   : 1.00
## 1st Qu.:25.75   1st Qu.: 24.50   1st Qu.:25.50   1st Qu.: 23.00
## Median :55.00   Median : 49.00   Median :58.00   Median : 51.00
## Mean   :50.72   Mean   : 49.73   Mean   :51.14   Mean   : 49.87
## 3rd Qu.:73.00   3rd Qu.: 73.50   3rd Qu.:74.00   3rd Qu.: 75.25
## Max.   :99.00   Max.   :100.00   Max.   :99.00   Max.   :100.00
## knowledge comprehension application group group.dummy
##  Min.   : 3.00   Min.   : 2.0   Min.   : 4.00   1:18   0:45
## 1st Qu.: 31.00   1st Qu.:22.0   1st Qu.: 30.75   2:28   1:55
## Median : 58.50   Median :47.5   Median : 53.00   3:26
## Mean   : 53.99   Mean   :47.0   Mean   : 53.00   4:28
## 3rd Qu.: 79.50   3rd Qu.:71.0   3rd Qu.: 80.50
## Max.   :100.00   Max.   :97.0   Max.   :100.00
## attention      time      interaction
##  Min.   : 2.00   Min.   : 2.00   Min.   : 1.00
## 1st Qu.: 26.00   1st Qu.: 29.75   1st Qu.: 30.00
## Median : 52.00   Median : 52.50   Median : 49.00
## Mean   : 52.55   Mean   : 53.03   Mean   : 52.04
## 3rd Qu.: 79.25   3rd Qu.: 79.25   3rd Qu.: 77.25
## Max.   :100.00   Max.   :100.00   Max.   :100.00
```

```
boxplot(d_ges$knowledge, d_ges$comprehension, d_ges$application, col=c('turquoise1','turquoise2','turquoise3'),names=c("application","comprehension","knowledge"), ylab="learning outcome in %")
```



```
plot(d_ges$learning_outcome,d_ges$prior_knowledge, ylab='prior knowledge in %', xlab='learning outcome in %')
```

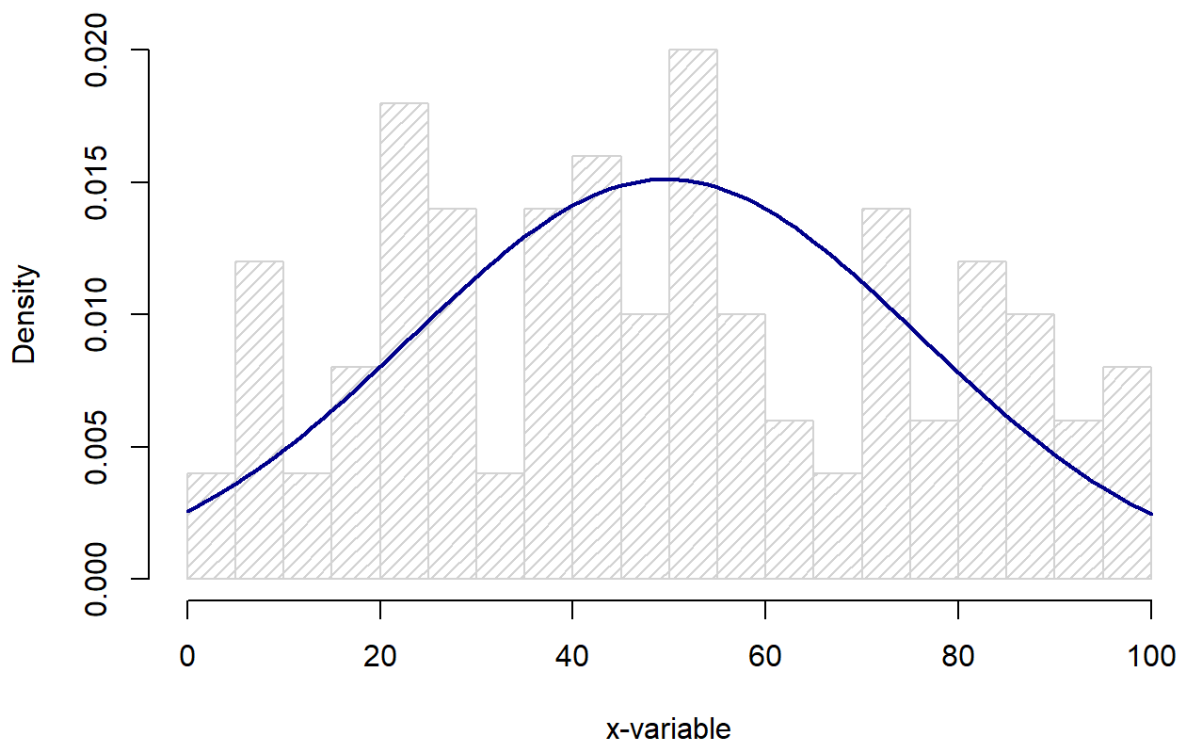


```

m<-mean(d_ges$age)
std<-sqrt(var(d_ges$age))
hist(d_ges$age, density=20, breaks=20, prob=TRUE,
      xlab="x-variable", ylim=c(0, 0.02),
      main="distribution of age with normal curve over histogram")
curve(dnorm(x, mean=m, sd=std),
      col="darkblue", lwd=2, add=TRUE, yaxt="n")

```

distribution of age with normal curve over histogram



```
skewness(d_ges$age)
```

```
## [1] 0.1268351
```

```
kurtosis(d_ges$age)
```

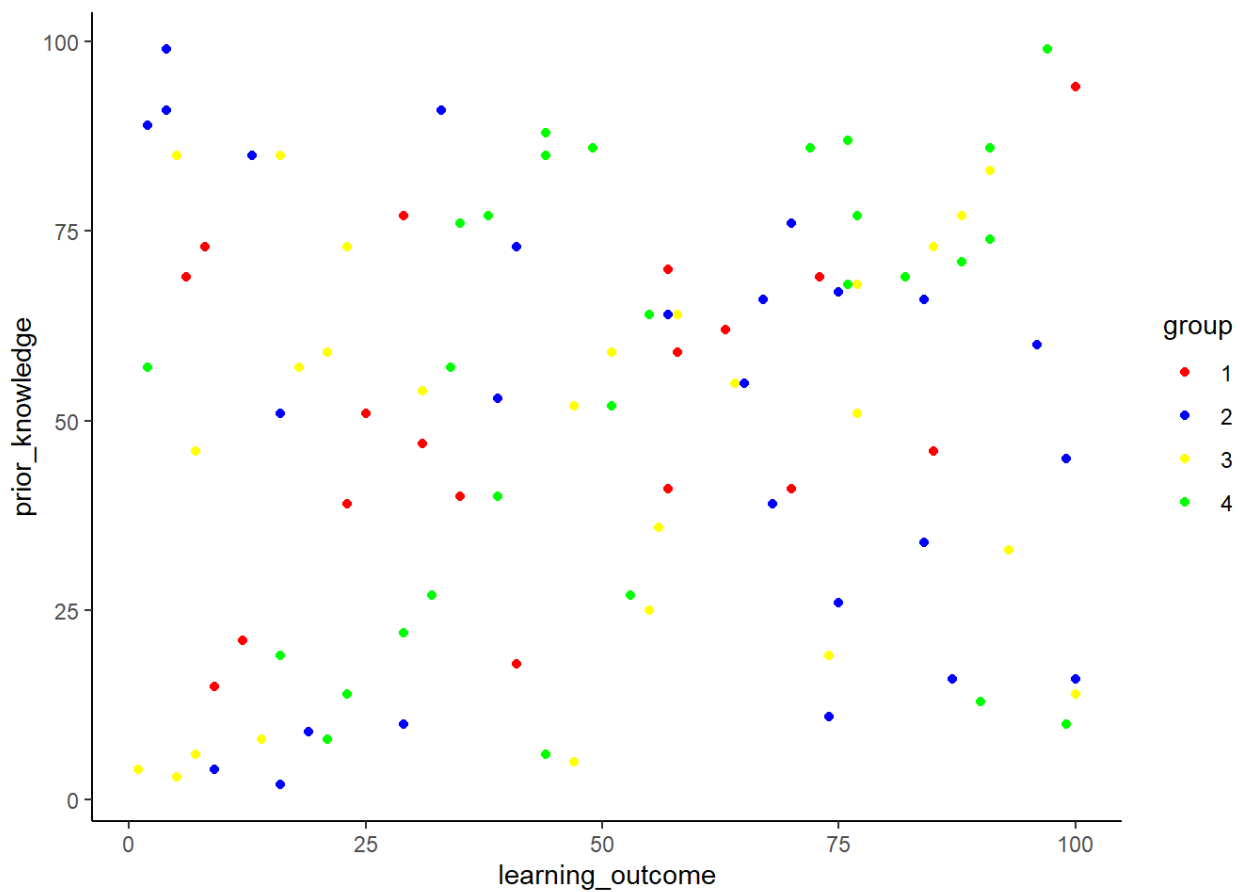
```
## [1] 1.958863
```

Plotting by group

```

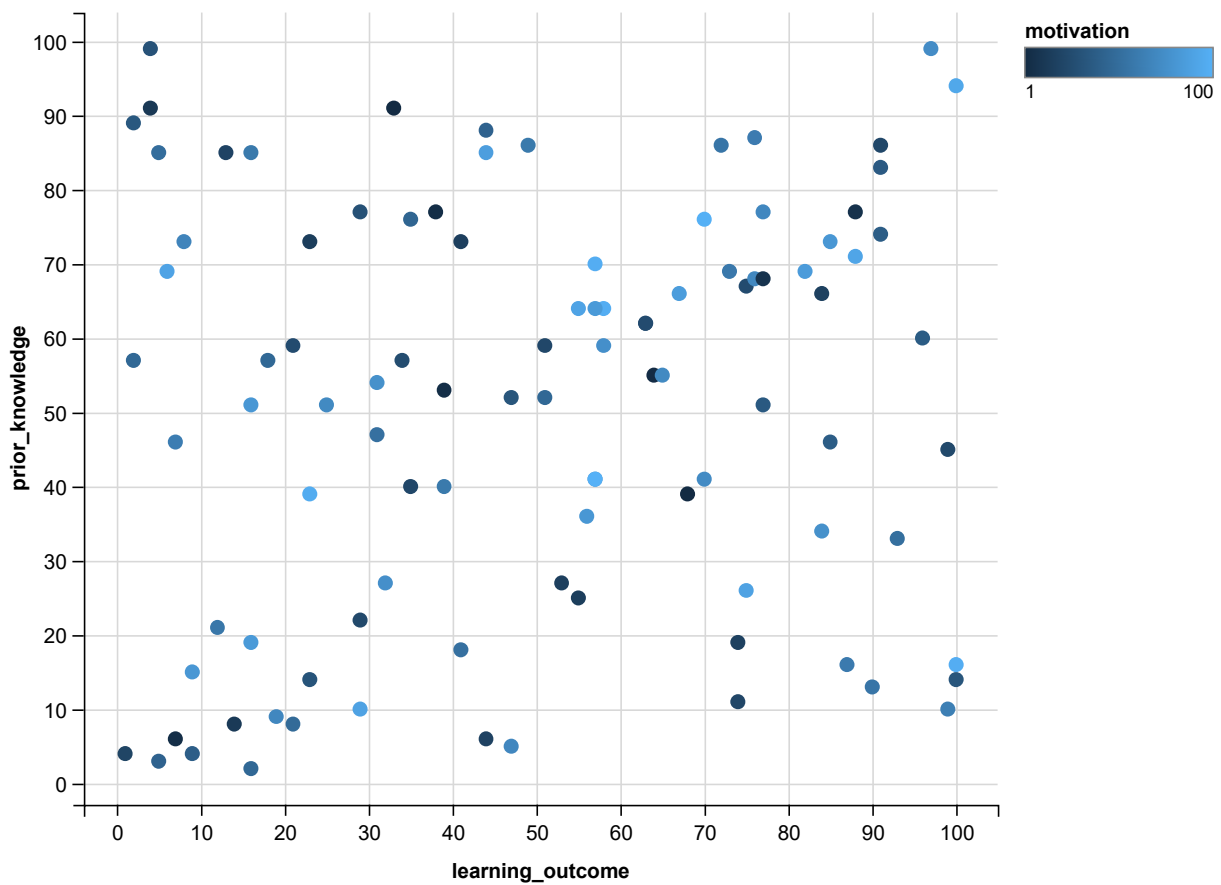
px1 <- ggplot(data = d_ges, aes(learning_outcome, prior_knowledge, color = group)) +
  geom_point() +
  scale_color_manual(values = c('red', 'blue', 'yellow', 'green'))
px2 <- px1 + theme_bw() + theme(panel.border = element_blank(), panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(), axis.line = element_line(colour = "black"))
px2

```

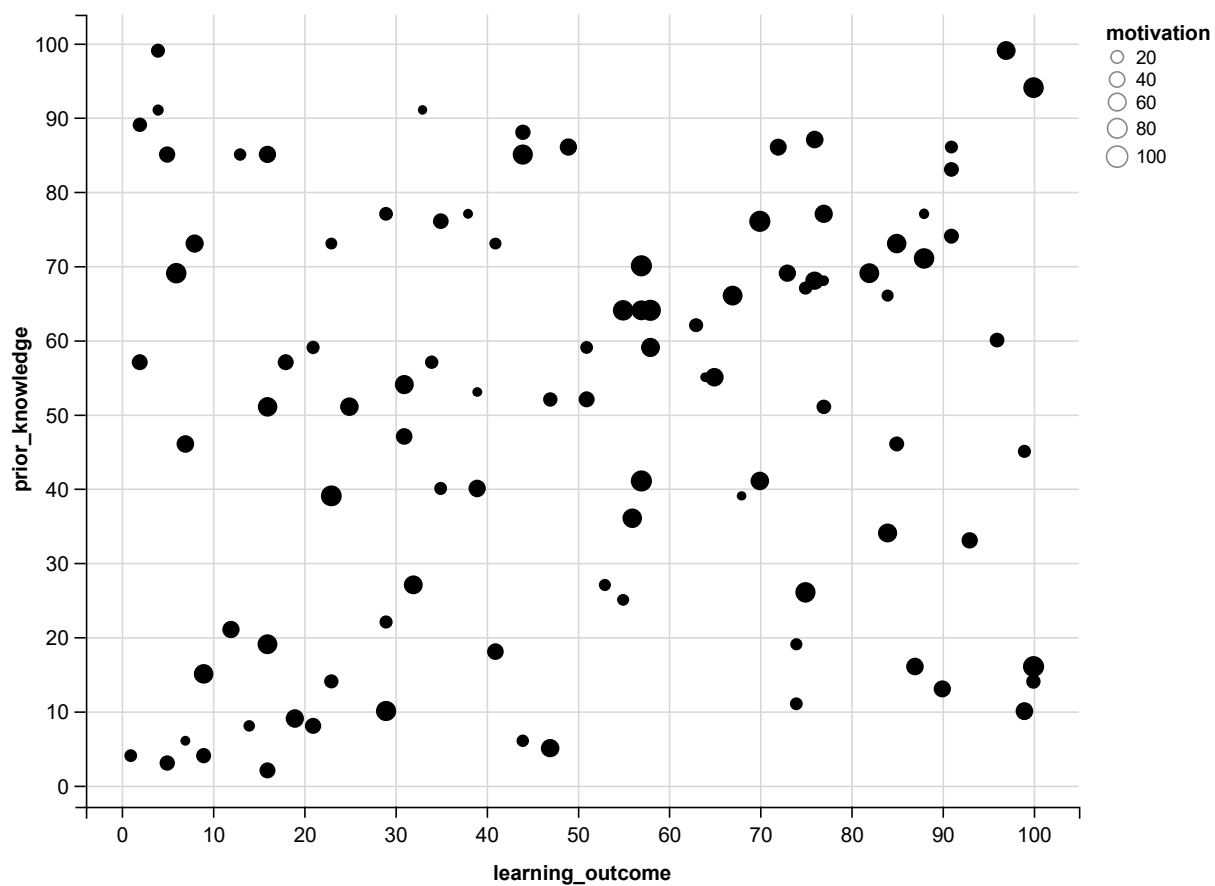


More versions of scatterplots using ggvis

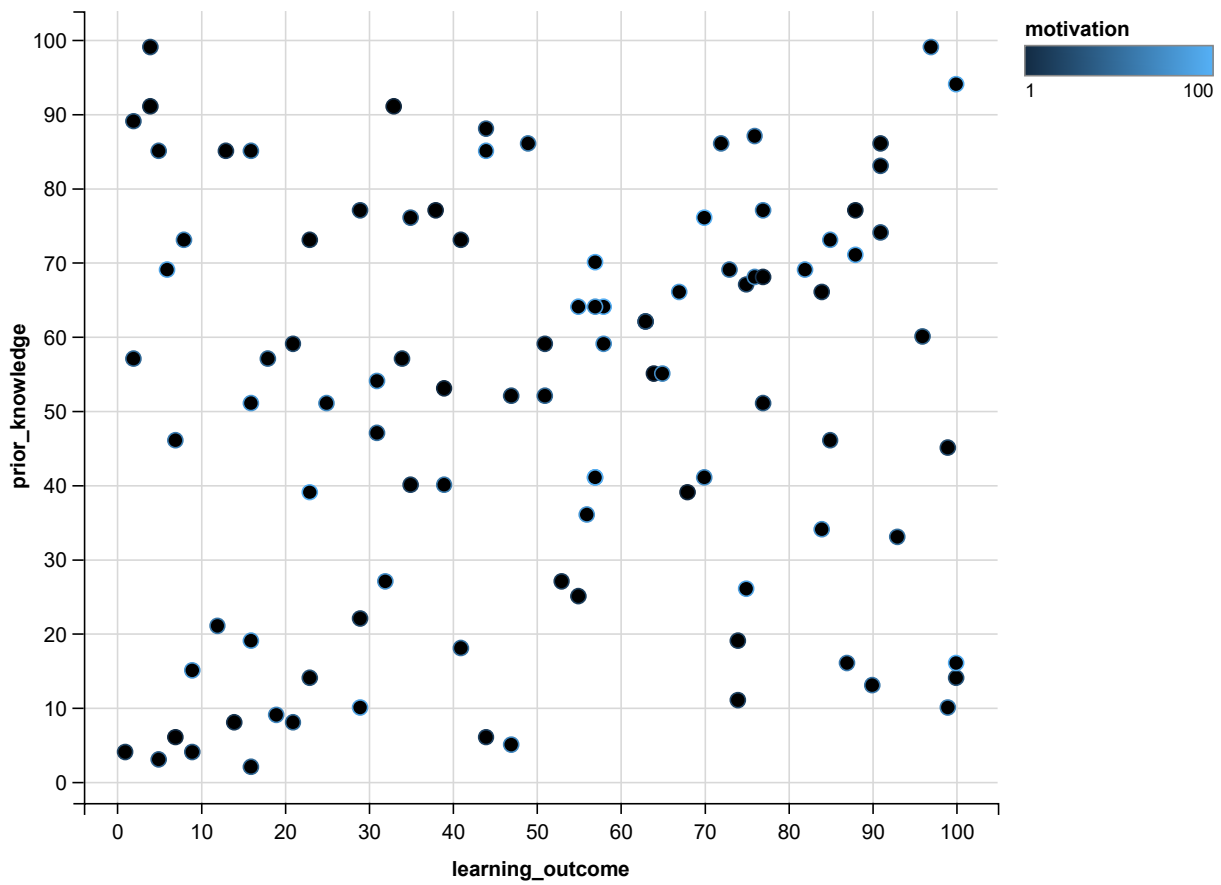
```
d_ges %>% ggvis(~learning_outcome,~prior_knowledge, fill = ~motivation) %>% layer_points
()
```



```
d_ges %>% ggvis(~learning_outcome,~prior_knowledge, size = ~motivation) %>% layer_points()
```

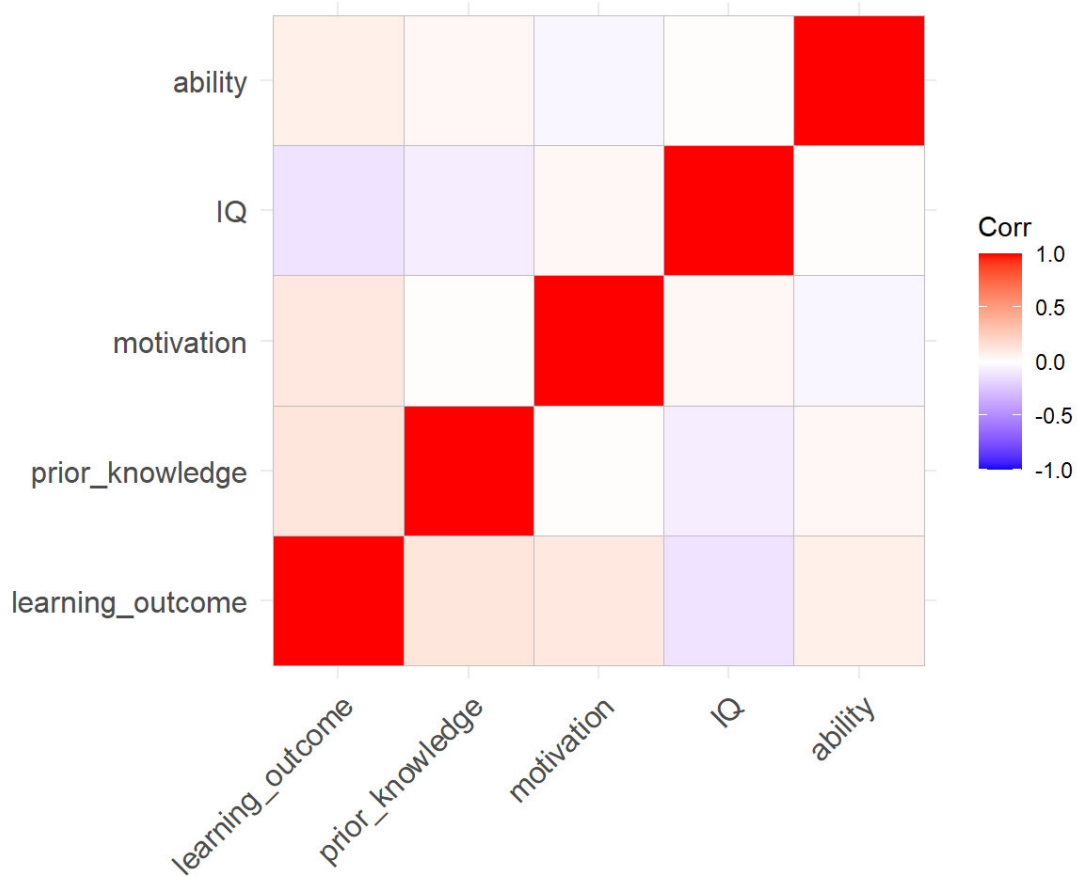


```
d_ges %>% ggvis(~learning_outcome,~prior_knowledge, stroke = ~motivation) %>% layer_points()
```



Plotting to explore correlations

```
d_cor <- subset(d_ges, select=c('learning_outcome','prior_knowledge','motivation','IQ','ability'))
ggcorrplot::ggcorrplot(cor(d_cor))
```



```
corrplot::corrplot(cor(d_cor))
```



Descriptive data

```
d_des <- subset(d_ges, select=c('group.dummy','learning_outcome','prior_knowledge','moti
vation','IQ'))
describeBy(d_des,d_des$group.dummy)
```

```
##
## Descriptive statistics by group
## group: 0
##
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
group.dummy*	1	45	1.00	0.00	1	1.00	0.00	1	1	0	NaN
learning_outcome	2	45	48.18	29.56	47	47.46	35.58	2	100	98	0.17
prior_knowledge	3	45	51.20	26.77	55	51.38	26.69	6	99	93	-0.21
motivation	4	45	49.40	28.20	56	49.92	37.06	2	96	94	-0.13
IQ	5	45	58.84	29.66	64	60.22	34.10	3	100	97	-0.35

```
##
```

	kurtosis	se
group.dummy*	NaN	0.00
learning_outcome	-1.12	4.41
prior_knowledge	-1.08	3.99
motivation	-1.34	4.20
IQ	-1.23	4.42

```
## -----
## group: 1
##
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
group.dummy*	1	55	2.00	0.00	2	2.00	0.00	2	2	0	NaN
learning_outcome	2	55	51.25	30.83	53	51.56	35.58	1	100	99	-0.11
prior_knowledge	3	55	50.33	28.83	52	51.07	37.06	2	94	92	-0.23
motivation	4	55	50.00	29.49	46	49.64	35.58	1	100	99	0.15
IQ	5	55	46.91	28.66	46	46.82	35.58	1	99	98	0.03

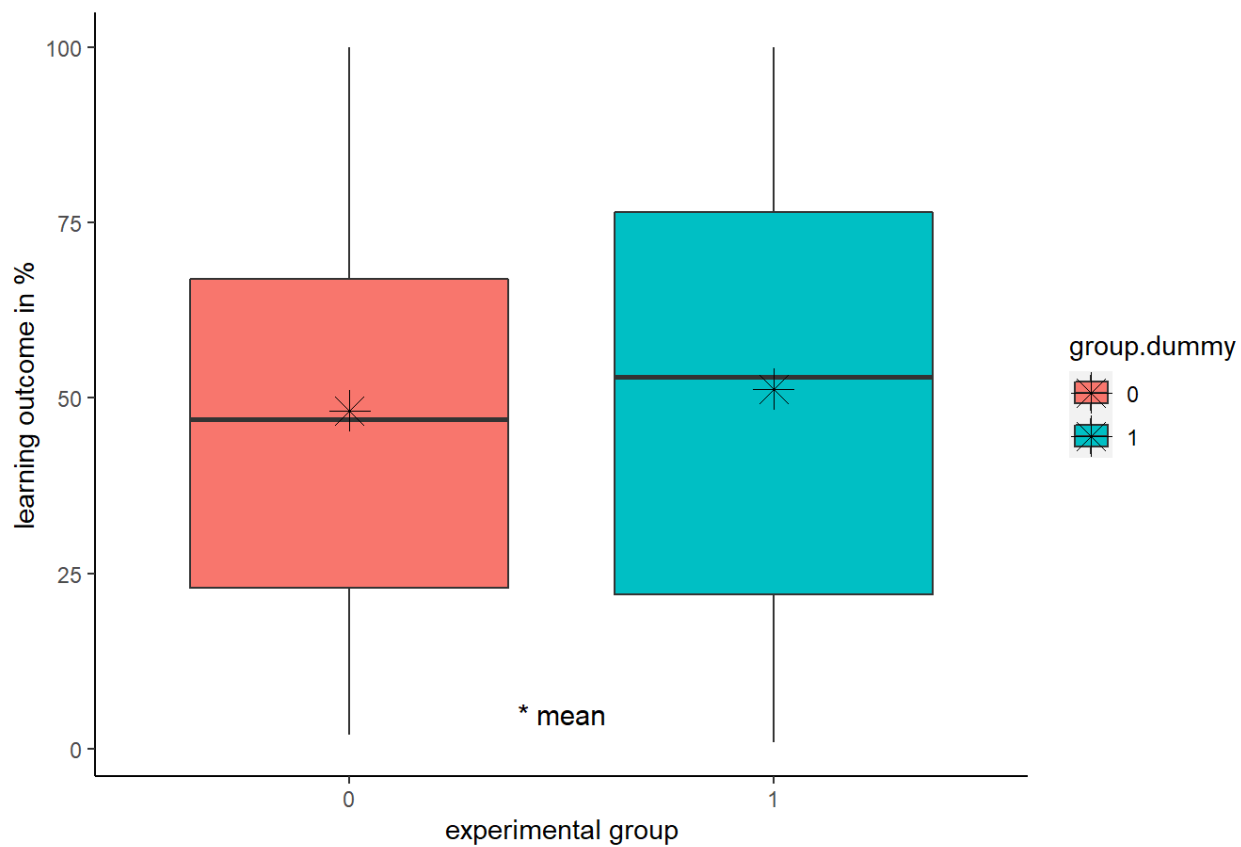
```
##
```

	kurtosis	se
group.dummy*	NaN	0.00
learning_outcome	-1.41	4.16
prior_knowledge	-1.31	3.89
motivation	-1.23	3.98
IQ	-1.27	3.86

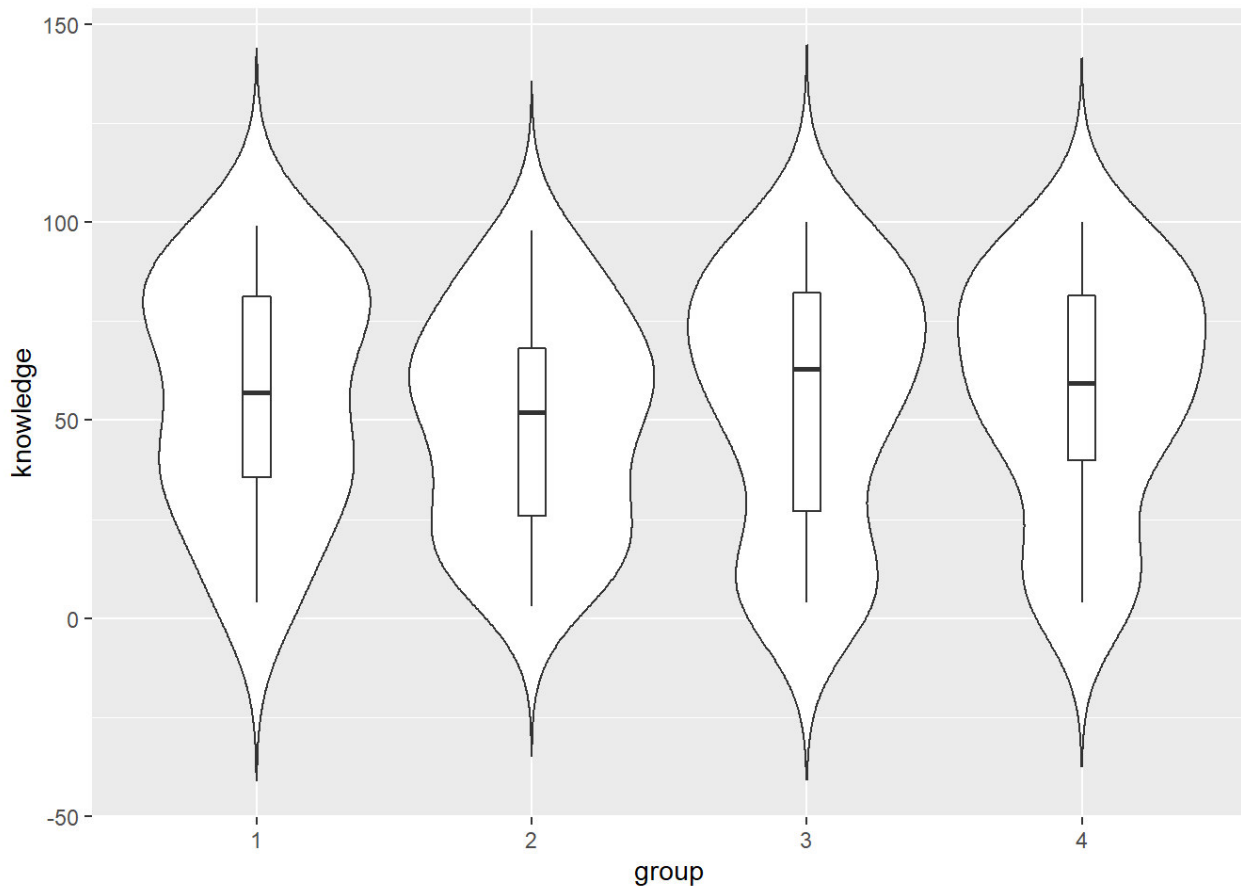
```
bp <- ggplot(d_des) +
  aes(x = group.dummy, y = learning_outcome, fill = group.dummy) +
  geom_boxplot() +
  stat_summary(fun=mean, colour="black", geom="point", shape=8, size=5) +
  scale_x_discrete(name = "experimental group") + scale_y_continuous(name = "learning outcome in %") +
  annotate("text", x=1.5, y=5, size=4, label="* mean") +
  ggtitle("Boxplot and mean of learning outcome for each experimental group")

bp1 <- bp + theme(panel.background = element_blank(), axis.line = element_line(colour = "black"))
bp1
```


Boxplot and mean of learning outcome for each experimental group



```
p <- ggplot(d_ges, aes(x=group, y=knowledge)) +  
  geom_violin(trim=FALSE)  
#p  
p1 <- p + geom_boxplot(width=0.1)  
p1
```



Detecting potential outliers

```
d_sel <- subset(d_ges, select = c('group', 'knowledge', 'comprehension', 'application', 'case'))

d_sel %>%
  group_by(group) %>%
  mahalanobis_distance(-case) %>%
  filter(is.outlier == TRUE) %>%
  as.data.frame()
```

```
## [1] knowledge      comprehension application      case      mahal.dist
## [6] is.outlier
## <0 Zeilen> (oder row.names mit Länge 0)
```

Checking assumptions for inferential statistics

Normal distribution can either be checked by the the Shapiro Wilk test

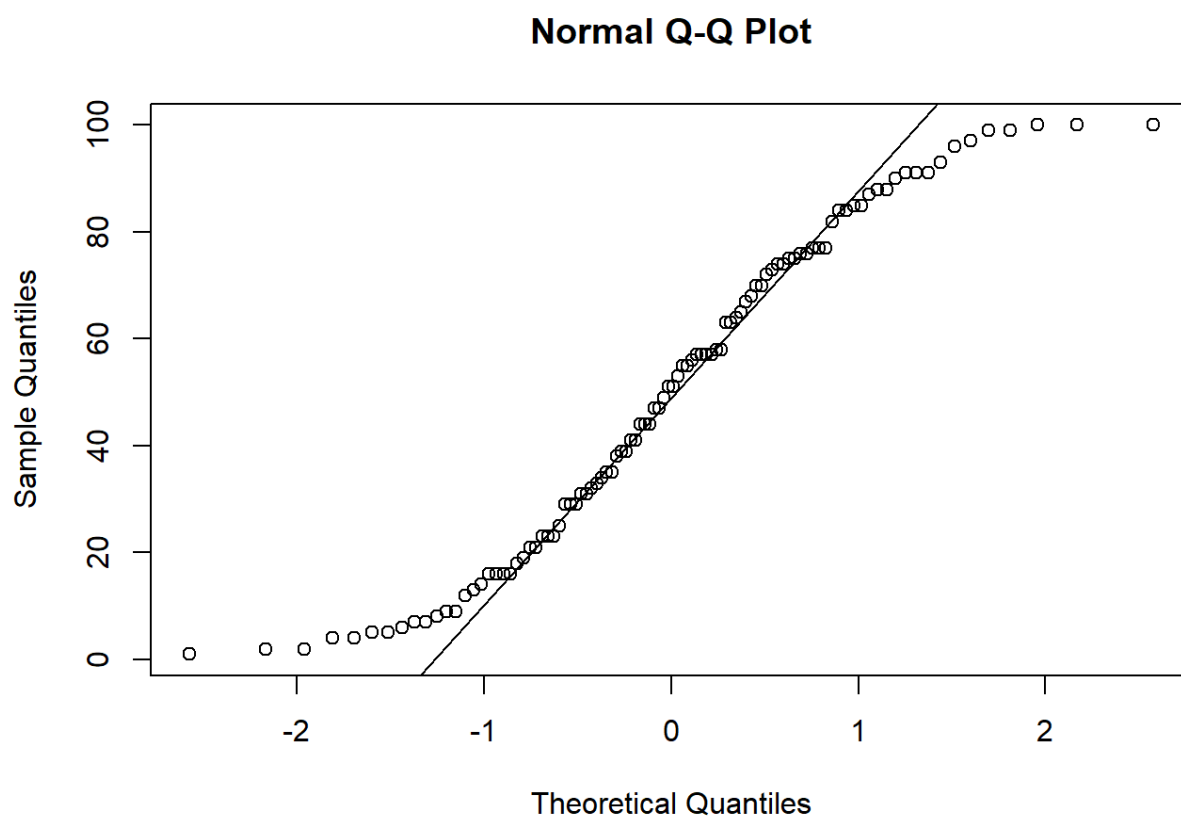
```
shapiro.test(d_ges$learning_outcome)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d_ges$learning_outcome
## W = 0.94698, p-value = 0.0005277
```

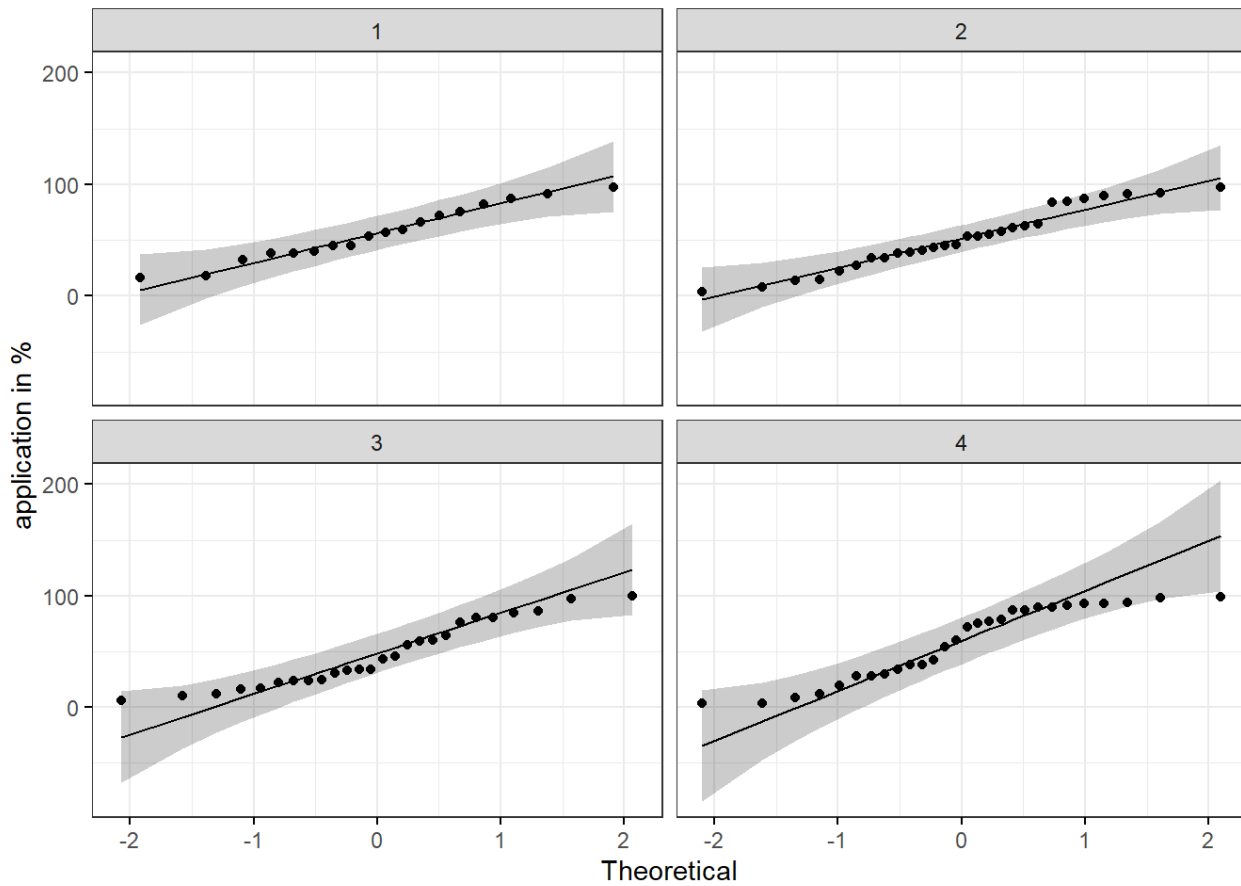
For $p > .05$ a normal distribution can be assumed

A visual inspection of the QQ-plot might deliver further insights.

```
#for all groups  
qqnorm(d_ges$learning_outcome)  
qqline(d_ges$learning_outcome)
```



```
#for different groups  
ggqqplot(d_sel, "application", facet.by = "group",  
          ylab = "application in %", ggtheme = theme_bw())
```



Visual inspection of normal distribution based on the qq plot: In case of a perfect normal distribution all points are on a straight line. In case the points form a curve instead of a straight line this is an indicator for skewness in the data.

Testing for multivariate normal distribution is required when more variables are included in the analysis.

```
result <- mvn(data=d_des, subset="group.dummy",mvnTest="hz")
result
```

```
## $multivariateNormality
## $multivariateNormality$`0`
##           Test           HZ      p value MVN
## 1 Henze-Zirkler 0.9737372 0.02988264 NO
##
## $multivariateNormality$`1`
##           Test           HZ      p value MVN
## 1 Henze-Zirkler 1.246329 0.0001571363 NO
##
##
## $univariateNormality
## $univariateNormality$`0`
##           Test           Variable Statistic      p value Normality
## 1 Anderson-Darling learning_outcome      0.4469      0.2688      YES
## 2 Anderson-Darling prior_knowledge      0.7751      0.0408      NO
## 3 Anderson-Darling motivation      0.8287      0.0299      NO
## 4 Anderson-Darling IQ      0.9902      0.0118      NO
##
## $univariateNormality$`1`
##           Test           Variable Statistic      p value Normality
## 1 Anderson-Darling learning_outcome      1.1606      0.0045      NO
## 2 Anderson-Darling prior_knowledge      1.0465      0.0087      NO
## 3 Anderson-Darling motivation      0.7212      0.0565      YES
## 4 Anderson-Darling IQ      0.7321      0.0531      YES
##
##
## $Descriptives
## $Descriptives$`0`
##           n      Mean  Std.Dev Median Min Max 25th 75th      Skew
## learning_outcome 45 48.17778 29.55756      47  2 100  23  67  0.1668663
## prior_knowledge  45 51.20000 26.76803      55  6  99  25  73 -0.2131473
## motivation       45 49.40000 28.20171      56  2  96  23  73 -0.1349775
## IQ               45 58.84444 29.66323      64  3 100  34  84 -0.3516555
##
##           Kurtosis
## learning_outcome -1.124425
## prior_knowledge  -1.084474
## motivation       -1.336459
## IQ               -1.234256
##
## $Descriptives$`1`
##           n      Mean  Std.Dev Median Min Max 25th 75th      Skew
## learning_outcome 55 51.25455 30.83241      53  1 100 22.0 76.5 -0.11067956
## prior_knowledge  55 50.32727 28.83224      52  2  94 26.5 75.0 -0.22731930
## motivation       55 50.00000 29.48634      46  1 100 25.0 75.5  0.15303607
## IQ               55 46.90909 28.66038      46  1  99 22.5 73.0  0.03153516
##
##           Kurtosis
## learning_outcome -1.407608
## prior_knowledge  -1.305880
## motivation       -1.229407
## IQ               -1.272762
```

When conducting analysis with more than one dependent variable multicollinearity need to be examined.

```
d_sel%>% cor_test(application, knowledge)
```

```
## # A tibble: 1 x 8
##   var1      var2      cor statistic      p conf.low conf.high method
##   <chr>    <chr>    <dbl>    <dbl> <dbl>    <dbl>    <dbl> <chr>
## 1 application knowledge -0.11      -1.10 0.276    -0.300    0.0884 Pearson
```

Another important assumption is the homogeneity of variances.

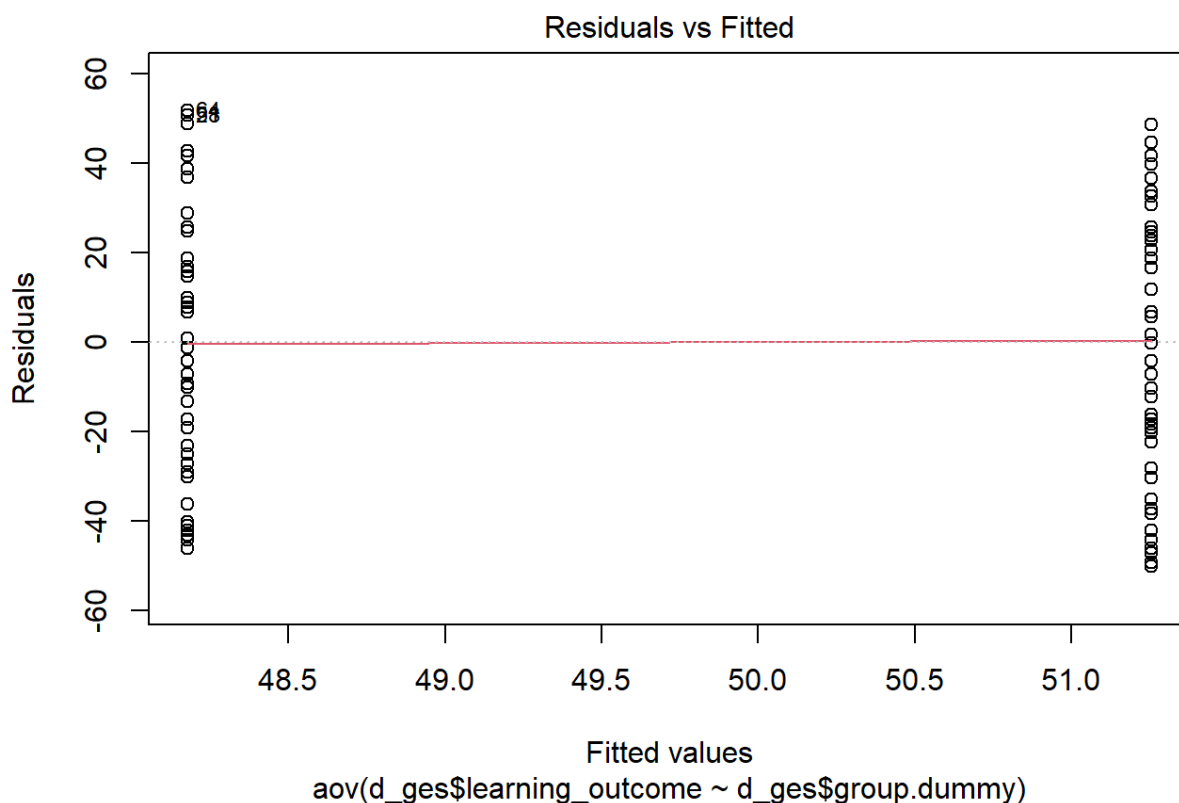
Analysing the homogeneity of variances by using the Bartlett's test

```
bartlett.test(d_ges$learning_outcome, d_ges$group.dummy)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: d_ges$learning_outcome and d_ges$group.dummy
## Bartlett's K-squared = 0.085307, df = 1, p-value = 0.7702
```

Additionally, the visual inspection of residuals can deliver further insights:

```
Mod.aov <- aov(d_ges$learning_outcome ~ d_ges$group.dummy)
plot(Mod.aov, 1)
```



Testing homogeneity of variance-covariance matrices with the Box's M Test

```
boxM(cbind(knowledge, comprehension, application) ~ group, data=d_ges)
```

```
##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data: Y
## Chi-Sq (approx.) = 11.769, df = 18, p-value = 0.8589
```

Inferential testing of hypothesis

Depending on the postulated hypothesis and the results from the testing of assumptions, different statistical test should be chosen. The most relevant statistical test related to this thesis are shown exemplary in the following.

To deduce whether two variables are significantly related the Pearson product-moment correlation can be used.

```
cor.test(d_ges$learning_outcome,d_ges$prior_knowledge)
```

```
##
## Pearson's product-moment correlation
##
## data: d_ges$learning_outcome and d_ges$prior_knowledge
## t = 1.3394, df = 98, p-value = 0.1835
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.06402401 0.32201776
## sample estimates:
## cor
## 0.1340802
```

When $p < .05$ a significant correlation is supported by the data.

Performing a t-test to deduce whether the groups differ significantly

```
t.test(d_ges$learning_outcome~d_ges$group.dummy, var.equal = TRUE, alternative = "two.sided")
```

```
##
## Two Sample t-test
##
## data: d_ges$learning_outcome by d_ges$group.dummy
## t = -0.50573, df = 98, p-value = 0.6142
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -15.149931 8.996396
## sample estimates:
## mean in group 0 mean in group 1
## 48.17778 51.25455
```

When $p < .05$ a significant difference between the groups is supported by the data.

In case that more than 2 experimental groups exist and significant differences should be examined on 1 dependent variable, an ANOVA can be performed.

```
Mod1 <- aov(learning_outcome~group, data=d_ges)
summary(Mod1)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1934    644.8   0.703  0.553
## Residuals  96  88075    917.4
```

```
pander(Mod1)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	1934	644.8	0.7028	0.5526
Residuals	96	88075	917.4	NA	NA

Post-hoc testing to examine which group differ significantly

```
ph <- TukeyHSD(aov(learning_outcome~group, data=d_ges))
ph
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = learning_outcome ~ group, data = d_ges)
##
## $group
##      diff      lwr      upr      p adj
## 2-1  8.198413 -15.72706 32.12389 0.8069636
## 3-1  3.132479 -21.15041 27.41537 0.9867061
## 4-1 11.841270 -12.08421 35.76674 0.5689647
## 3-2 -5.065934 -26.63484 16.50297 0.9273641
## 4-2  3.642857 -17.52286 24.80857 0.9694769
## 4-3  8.708791 -12.86011 30.27770 0.7171195
```

Additionally, further independent variables or covariates might be added and an ANCOVA might be performed

```
Mod2 <- aov(learning_outcome~group + motivation, data=d_ges)
summary(Mod2)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1934    644.8   0.707  0.550
## motivation  1   1479   1479.5   1.623  0.206
## Residuals  95  86595    911.5
```

```
pander(Mod2)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	1934	644.8	0.7074	0.5499
motivation	1	1479	1479	1.623	0.2058
Residuals	95	86595	911.5	NA	NA

Potential interaction effects might be added

```
Mod3 <- aov(learning_outcome~group*motivation, data=d_ges)
summary(Mod3)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3   1934    644.8    0.695  0.558
## motivation      1   1479   1479.5    1.594  0.210
## group:motivation 3   1190    396.7    0.427  0.734
## Residuals      92  85405    928.3
```

```
pander(Mod3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	1934	644.8	0.6946	0.5577
motivation	1	1479	1479	1.594	0.21
group:motivation	3	1190	396.7	0.4273	0.7339
Residuals	92	85405	928.3	NA	NA

To test more specific hypothesis: contrast analysis can be used.

```
d_ges$group <- relevel(d_ges$group, ref="1")
modell1.contrref <- lm(learning_outcome ~ group, data=d_ges)
summary(modell1.contrref)
```

```
##
## Call:
## lm(formula = learning_outcome ~ group, data = d_ges)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.286 -25.754   0.069  24.157  56.556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   43.444      7.139   6.085 2.38e-08 ***
## group2         8.198      9.151   0.896   0.373
## group3         3.132      9.287   0.337   0.737
## group4        11.841      9.151   1.294   0.199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.29 on 96 degrees of freedom
## Multiple R-squared:  0.02149,    Adjusted R-squared:  -0.009088
## F-statistic: 0.7028 on 3 and 96 DF,  p-value: 0.5526
```

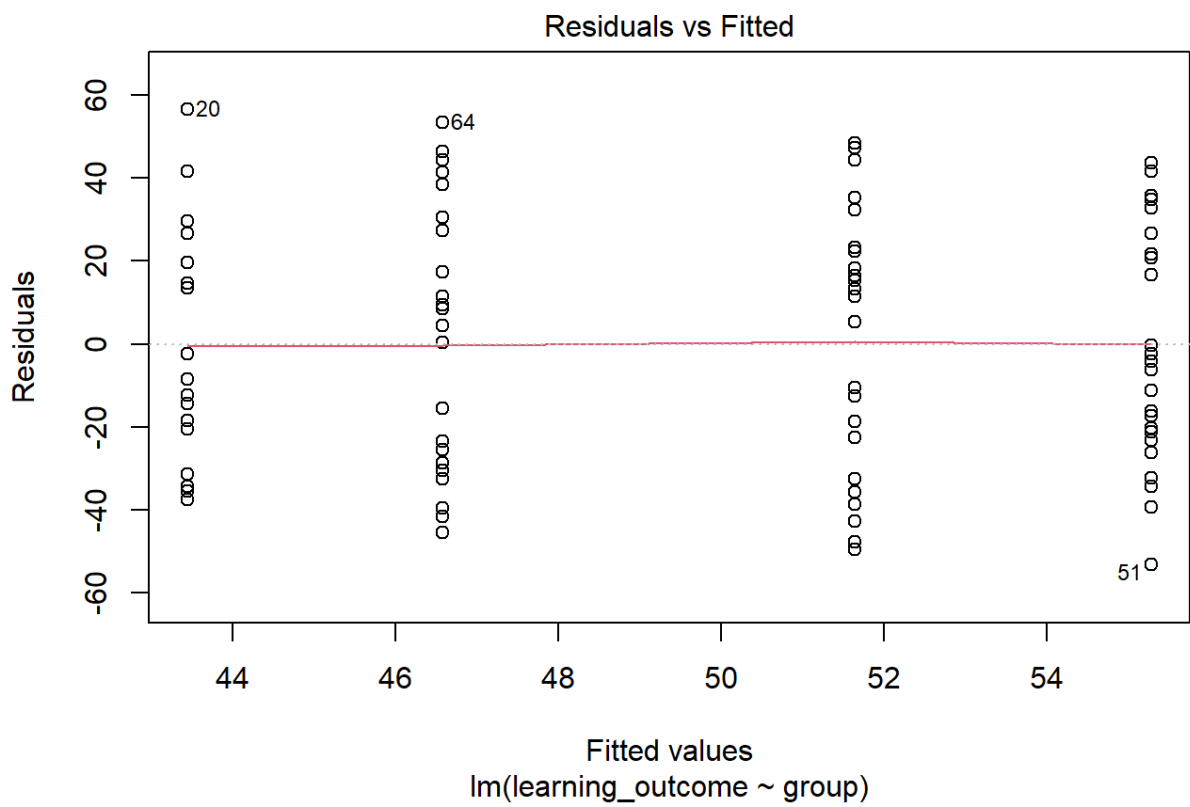
```
pander(modell1.contrref)
```

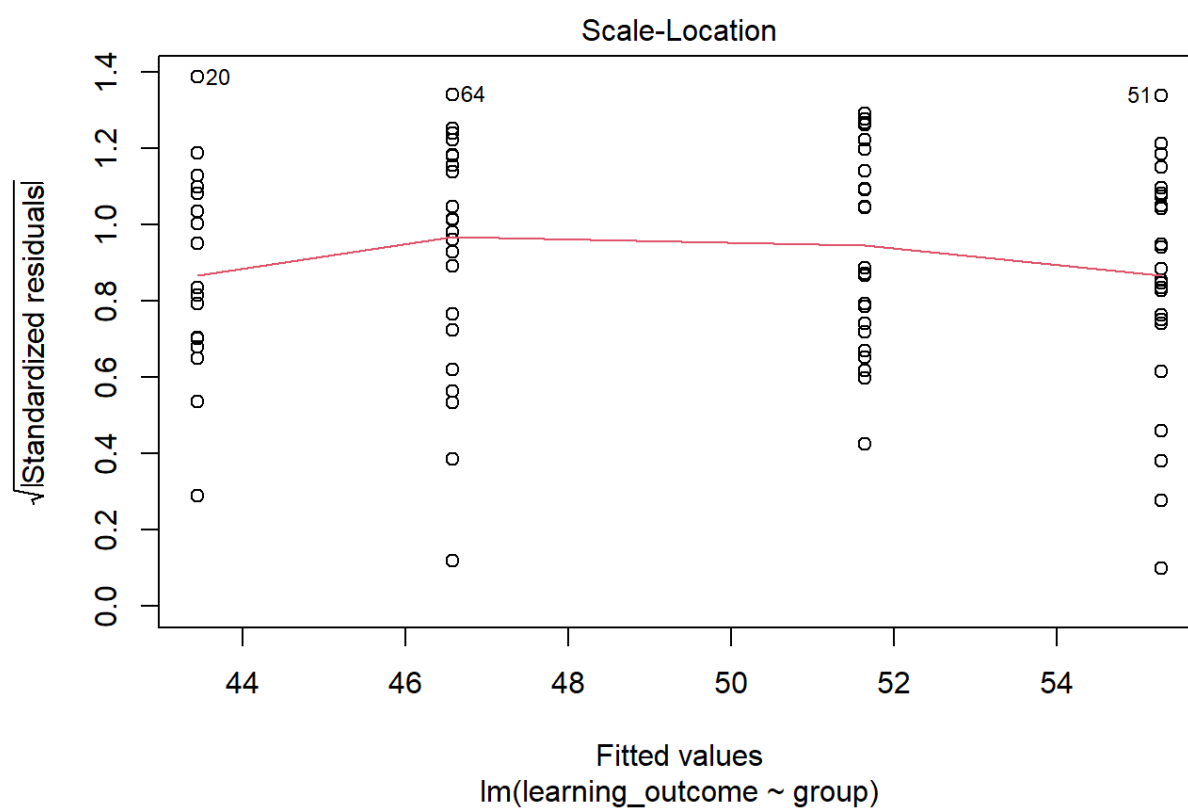
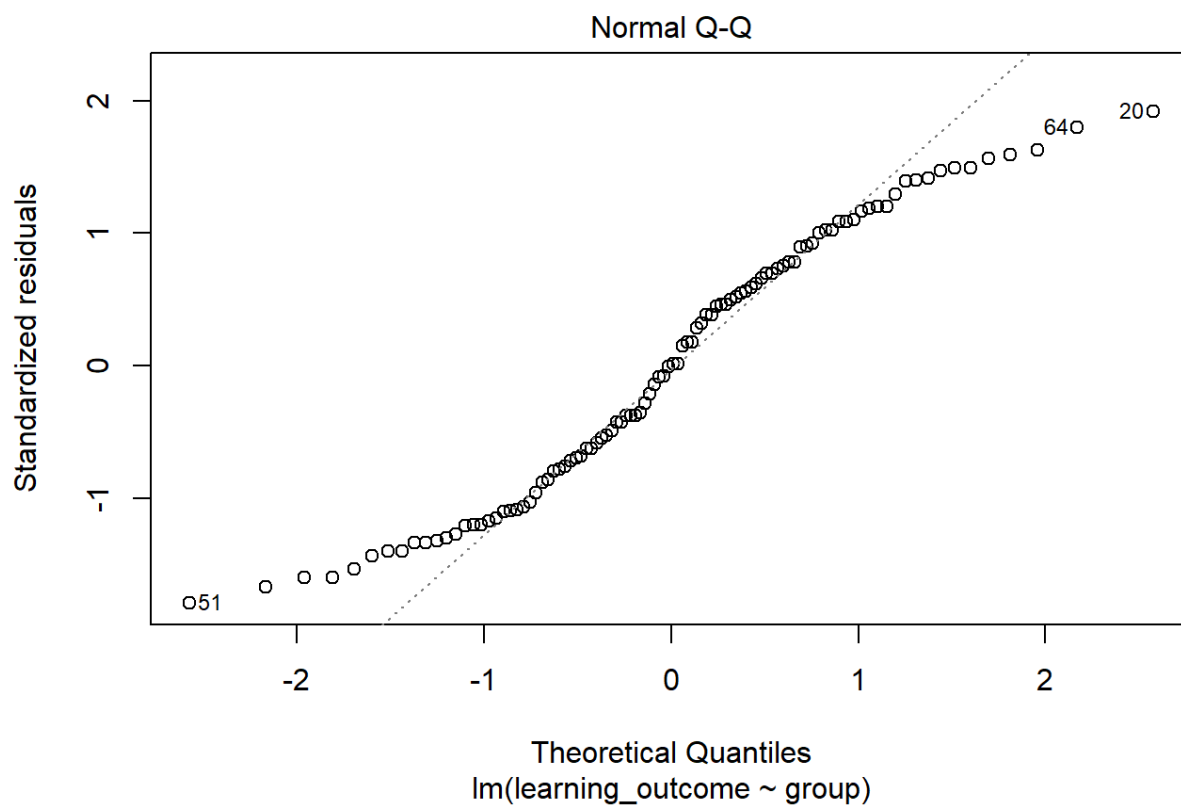
Fitting linear model: learning_outcome ~ group

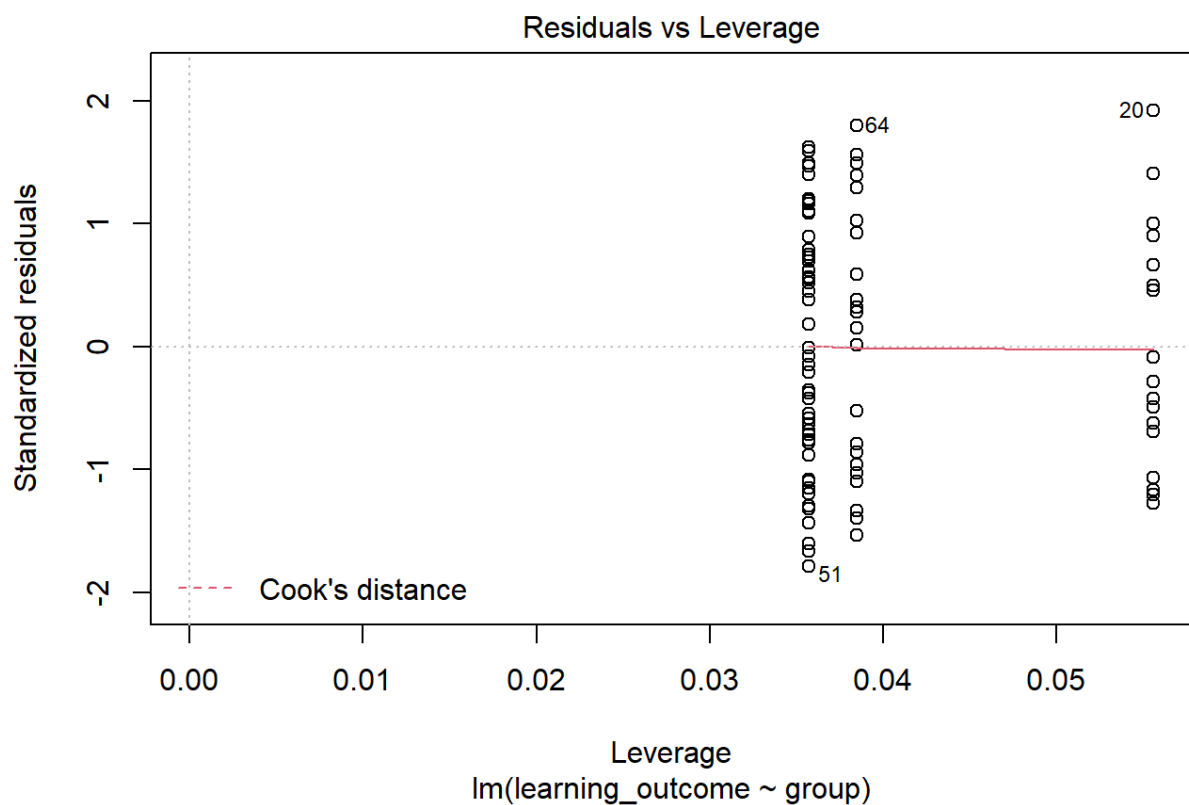
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	43.44	7.139	6.085	2.376e-08
group2	8.198	9.151	0.8959	0.3725
group3	3.132	9.287	0.3373	0.7366
group4	11.84	9.151	1.294	0.1988

After conducting the analysis, the results can be analyzed in more detail by additional plots for outlier detection.

```
#plotting residual, qq plot and leverage
plot(modell1.contrref )
```







When interpreting complex interactions, a graphical inspection of the given pattern is very helpful.

```
d_ges1$group.dummy <- as.numeric(d_ges1$group.dummy)
Mod4<- lm(learning_outcome~group.dummy+IQ + (group.dummy*IQ), data=d_ges1)
pander(summary(Mod4))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.691	2.705	-2.474	0.01636
group.dummy	5.635	1.78	3.165	0.002489
IQ	1.065	0.3023	3.523	0.0008482
group.dummy:IQ	-0.6089	0.1978	-3.078	0.003203

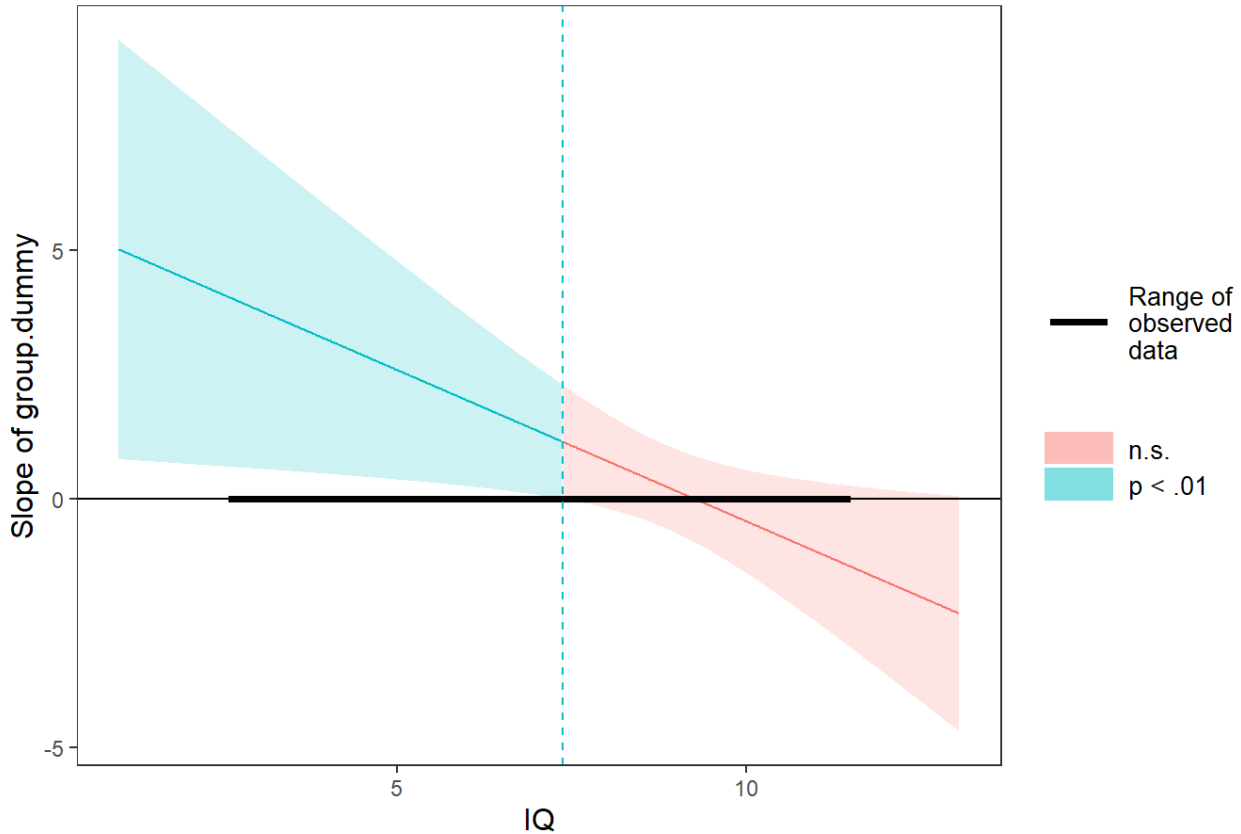
Fitting linear model: learning_outcome ~ group.dummy + IQ + (group.dummy * IQ)

Observations	Residual Std. Error	R^2	Adjusted R^2
61	1.222	0.1948	0.1524

```
SS_Res <- sim_slopes(Mod4, pred = group.dummy, modx = IQ, johnson_neyman = TRUE)
johnson_neyman(Mod4, pred = group.dummy, modx = IQ, alpha = 0.01)
```

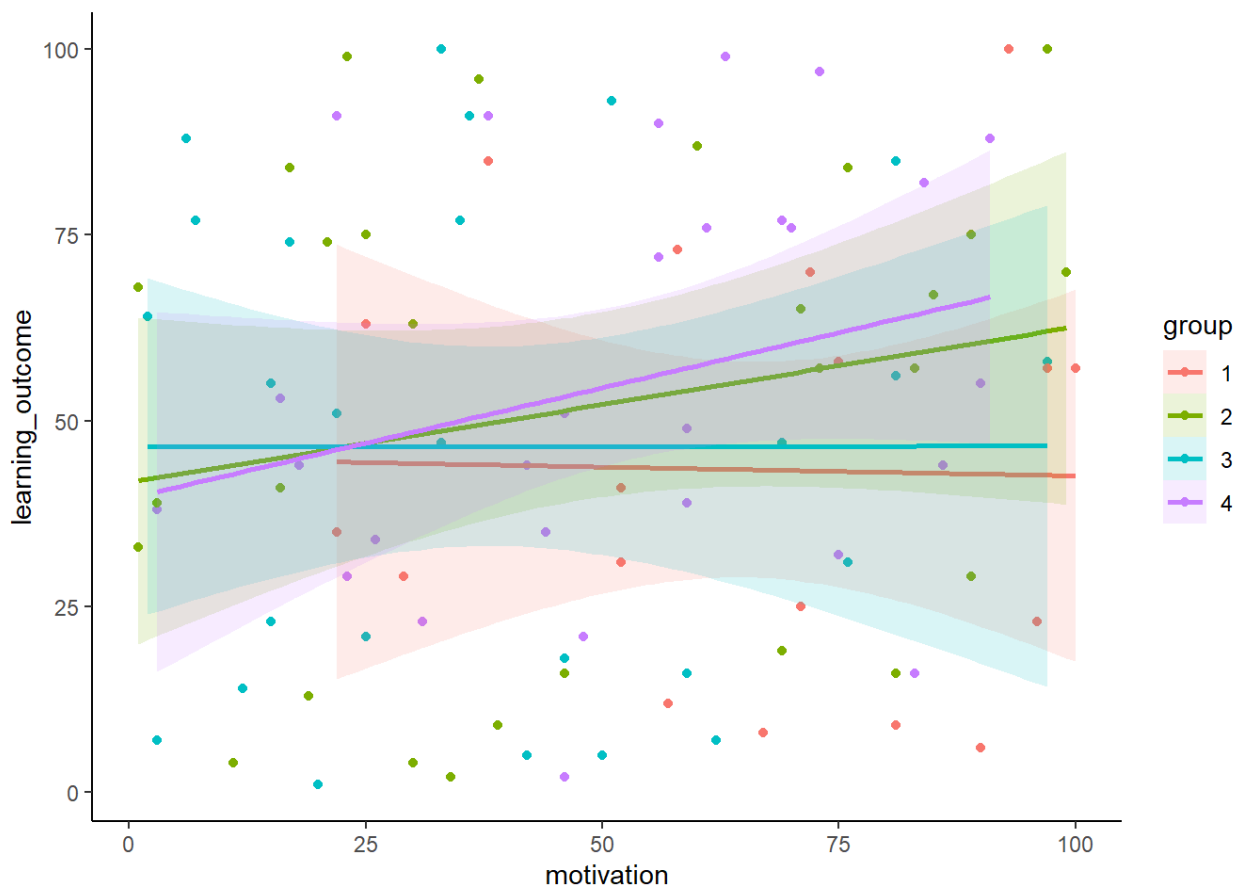
```
## JOHNSON-NEYMAN INTERVAL
##
## When IQ is OUTSIDE the interval [7.37, 13.51], the slope of group.dummy is
## p < .01.
##
## Note: The range of observed values of IQ is [2.63, 11.43]
```

Johnson-Neyman plot



```
pci <- ggplot(d_ges, aes(x = motivation, y = learning_outcome, color = group) ) +
  geom_point() +
  geom_smooth(formula = y ~ x, method = "lm", alpha = .15, aes(fill = group))

pci1 <- pci + theme_bw() + theme(panel.border = element_blank(), panel.grid.major = elem
ent_blank(),
panel.grid.minor = element_blank(), axis.line = element_line(colour = "black"))
pci1
```



Conducting a MANOVA in case more than one dependent variable is tested.

```
res.man <- manova(cbind(knowledge, comprehension, application) ~ group, d_ges)
summary(res.man)
```

```
##           Df  Pillai approx F num Df den Df Pr(>F)
## group      3  0.040167  0.43426      9  288  0.916
## Residuals 96
```

Post-hoc testing for investigating specific group differences

```
summary.aov(res.man)
```

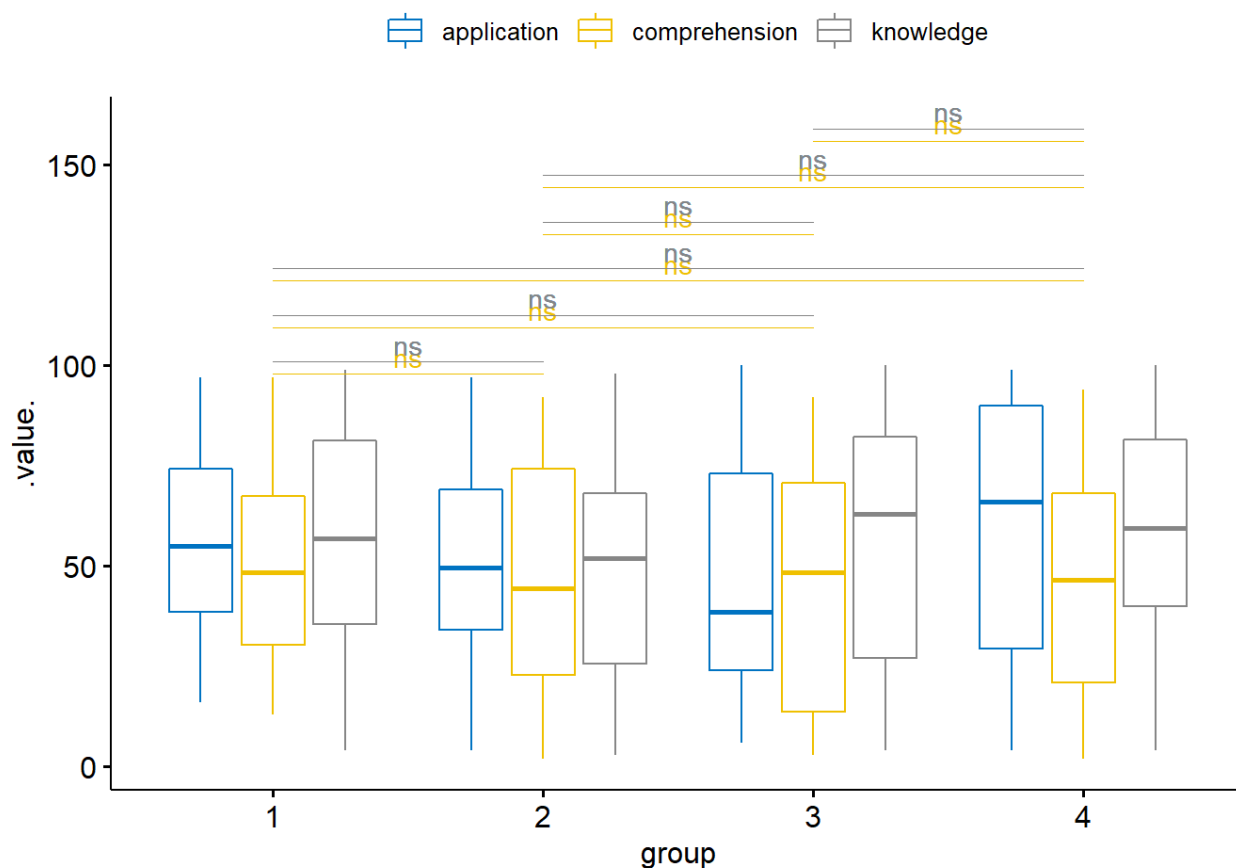
```
## Response knowledge :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1002   334.02  0.3795  0.768
## Residuals  96  84501   880.22
##
## Response comprehension :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3    173    57.56  0.0702  0.9757
## Residuals  96  78731   820.12
##
## Response application :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3    1921   640.28  0.7637  0.5172
## Residuals  96  80485   838.39
```

```
ph2 <- d_sel %>%
  gather(key = "variables", value = "value", application, comprehension, knowledge) %>%
  group_by(variables) %>%
  games_howell_test(value ~ group) %>%
  select(-estimate, -conf.low, -conf.high) # Remove details
ph2
```

```
## # A tibble: 18 x 6
##   variables      .y.  group1 group2 p.adj p.adj.signif
##   <chr>         <chr> <chr>  <chr>  <dbl> <chr>
## 1 application  value 1      2      0.931 ns
## 2 application  value 1      3      0.665 ns
## 3 application  value 1      4      0.996 ns
## 4 application  value 2      3      0.932 ns
## 5 application  value 2      4      0.85  ns
## 6 application  value 3      4      0.553 ns
## 7 comprehension value 1      2      0.99  ns
## 8 comprehension value 1      3      0.964 ns
## 9 comprehension value 1      4      0.982 ns
## 10 comprehension value 2      3      0.998 ns
## 11 comprehension value 2      4      1      ns
## 12 comprehension value 3      4      0.999 ns
## 13 knowledge    value 1      2      0.85  ns
## 14 knowledge    value 1      3      0.996 ns
## 15 knowledge    value 1      4      1      ns
## 16 knowledge    value 2      3      0.931 ns
## 17 knowledge    value 2      4      0.731 ns
## 18 knowledge    value 3      4      0.984 ns
```

Visualizing the results of post-hoc testing.

```
ph2 <- ph2 %>% add_xy_position(x = "group")
test.label <- create_test_label(
  description = "MANOVA", statistic.text = quote(italic("F")),
  statistic = 71.83, p = "<0.0001", parameter = "4,294",
  type = "expression", detailed = TRUE
)
ggboxplot(
  d_sel, x = "group", y = c("application", "comprehension", "knowledge"),
  merge = TRUE, palette = "jco"
) +
  stat_pvalue_manual(
    ph2, hide.ns = FALSE, tip.length = 0,
    step.increase = 0.1, step.group.by = "variables",
    color = "variables"
  )
```



Similar to extending the ANOVA, the MANOVA can be extended by including interactions.

```
res.man2 <- manova(cbind(knowledge, comprehension, application) ~ group*motivation, d_ge
s)
summary(res.man2)
```

```
##              Df    Pillai approx F num Df den Df Pr(>F)
## group          3 0.041001  0.42493      9   276 0.9212
## motivation      1 0.040901  1.27935      3    90 0.2863
## group:motivation 3 0.040709  0.42186      9   276 0.9229
## Residuals      92
```

```
summary.aov(res.man)
```

```
## Response knowledge :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3   1002   334.02  0.3795  0.768
## Residuals      96  84501   880.22
##
## Response comprehension :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3    173    57.56  0.0702 0.9757
## Residuals      96  78731   820.12
##
## Response application :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3   1921   640.28  0.7637 0.5172
## Residuals      96  80485   838.39
```

Potential covariates can be added to conduct a MANCOVA.


```
res.man2 <- manova(cbind(knowledge, comprehension, application) ~ group +prior_knowledg
e, d_ges)
summary(res.man2)
```

```
##              Df    Pillai approx F num Df den Df Pr(>F)
## group          3 0.040227  0.43038      9   285 0.9182
## prior_knowledge 1 0.024353  0.77379      3    93 0.5115
## Residuals      95
```

```
summary.aov(res.man)
```

```
## Response knowledge :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3   1002   334.02  0.3795  0.768
## Residuals      96  84501   880.22
##
## Response comprehension :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3    173    57.56  0.0702 0.9757
## Residuals      96  78731   820.12
##
## Response application :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group          3   1921   640.28  0.7637 0.5172
## Residuals      96  80485   838.39
```

Again, interactions can be added into the MANCOVA.

```
res.man2 <- manova(cbind(knowledge, comprehension, application) ~ group*motivation + pri
or_knowledge, d_ges)
summary(res.man2)
```

```
##              Df    Pillai approx F num Df den Df Pr(>F)
## group          3 0.041073  0.42106      9   273 0.9234
## motivation      1 0.040980  1.26768      3    89 0.2904
## prior_knowledge 1 0.025666  0.78147      3    89 0.5074
## group:motivation 3 0.035378  0.36197      9   273 0.9522
## Residuals      91
```

```
summary.aov(res.man)
```

```
## Response knowledge :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1002   334.02   0.3795  0.768
## Residuals  96  84501   880.22
##
## Response comprehension :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3    173    57.56   0.0702 0.9757
## Residuals  96  78731   820.12
##
## Response application :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1921   640.28   0.7637 0.5172
## Residuals  96  80485   838.39
```

For more directed and more specific hypothesis, contrasts can be included in the MANOVA.

```
outcome <- cbind(d_ges$knowledge, d_ges$comprehension, d_ges$application)
CM1 <- manova(outcome ~ group, data=d_ges)

summary(CM1, intercept=TRUE)
```

```
##           Df Pillai approx F num Df den Df Pr(>F)
## (Intercept)  1 0.91315    329.43      3    94 <2e-16 ***
## group        3 0.04017      0.43      9   288  0.916
## Residuals    96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary.aov(CM1)
```

```
## Response 1 :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1002   334.02   0.3795  0.768
## Residuals  96  84501   880.22
##
## Response 2 :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3    173    57.56   0.0702 0.9757
## Residuals  96  78731   820.12
##
## Response 3 :
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      3   1921   640.28   0.7637 0.5172
## Residuals  96  80485   838.39
```

```
contrast1<- c(1, 0, 0, 0)
contrast2<- c(0, 1, 0, 0)
contrast3<- c(0, 0, 1, 0)
contrasts(d_ges$group) <- cbind(contrast1,contrast2,contrast3)

CM2 <- lm(outcome ~ group, data=d_ges)
CM2
```

```
##  
## Call:  
## lm(formula = outcome ~ group, data = d_ges)  
##  
## Coefficients:  
##           [,1]      [,2]      [,3]  
## (Intercept)  57.1429  46.5714  58.0714  
## groupcontrast1 -0.8651   2.9841  -1.9048  
## groupcontrast2 -7.8929   0.5000  -6.5357  
## groupcontrast3 -3.0275  -0.9560 -11.1484
```


List of Figures

2.1. Simplified version of the Cognitive Affective Model of Immersive Learning (Makransky & Peterson, 2021, p. 7)	10
2.2. Adapted version of the Tetrahedral Model of Memory Experiments by Jenkins (1979)	14
2.3. Simplified version of the Cognitive Theory of Multimedia Learning (Mayer, 2020, p. 40)	16
2.4. Simplified version of the Integrated Model of Text and Picture Comprehension (Schnotz & Bannert, 2003, p. 145)	17
3.1. Underlying cognitive learning processes while learning in virtual reality included in Extended Model of Immersive Learning in Virtual Reality	32
3.2. Internal and external support approaches included in Extended Model of Immersive Learning in Virtual Reality	33
3.3. Learner's characteristics included in Extended Model of Immersive Learning in Virtual Reality	34
3.4. Extended Model of Immersive Learning in Virtual Reality (EMIL VR)	37
4.1. Advanced organizer based on the Extended Model of Immersive Learning in Virtual Reality and the three experimental studies	43
4.2. Procedure of Study 1	45
4.3. Procedure of Study 2 (Vogt, Babel, et al., 2021b, p.7)	48
4.4. Procedure of Study 3	51
5.1. Adapted version of the Extended Model of Immersive Learning in Virtual Reality	57
5.2. Adapted version of the Extended Model of Immersive Learning in Virtual Reality	63
1. Johnson-Neyman interval displaying the missing moderating effect of intrinsic motivation on application level	150
2. Johnson-Neyman interval displaying the moderating effect of intrinsic motivation on comprehension level	151
3. Relationship of prior knowledge and overall learning outcome	152

List of Acronyms

CAMIL	Cognitive Affective Model of Immersive Learning
CATLM	Cognitive-Affective Theory of Learning with Media
CTML	Cognitive Theory of Multimedia Learning
EMIL	Extended Model of Immersive Learning
F1	learning setting
F2	learning material
F3	learner's characteristics
F4	assessment of learning outcome
H	Hypothesis
IMTPC	Integrated Model of Text and Picture Comprehension
IQ	Intelligence quotient
M	Mean
Max	Maximum
Min	Minimum
N	sample
n	subsample
Q	Question
RQ	Research Question
SD	Standard deviation
SAK	Structural Assessment of Knowledge
VR	Virtual Reality
VRLE	Virtual Reality Learning Environment

Declaration

I hereby declare that I have prepared this thesis independently and have not used any sources or aids other than those indicated, and that I have clearly identified any passages taken verbatim or in substance, and that I have complied with the currently valid statutes of the University of Ulm for the safeguarding of good scientific practice (§6 Abs. 2 Satz 2 Promotionsordnung). Language editing advice was given by Raven Rinas.