Investigating Representation of Text and Audio in Educational VR using Learning Outcomes and EEG

Sarune Baceviciute, Aske Mottelson, Thomas Terkildsen, and Guido Makransky
Department of Psychology, University of Copenhagen
Copenhagen, Denmark
{sb, amot, tst, gm}@psy.ku.dk

ABSTRACT
This paper reports findings from a between-subjects experiment that investigates how different learning content representations in virtual environments (VE) affect the process and outcomes of learning. Seventy-eight participants were subjected to an immersive virtual reality (VR) application, where they received identical instructional information, rendered in three different formats: as text in an overlay interface, as text embedded semantically in a virtual book, or as audio. Learning outcome measures, self-reports, and an electroencephalogram (EEG) were used to compare conditions. Results show that reading was superior to listening for the learning outcomes of retention, self-efficacy, and extraneous attention. Reading text from a virtual book was reported to be less cognitively demanding, compared to reading from an overlay interface. EEG analyses show significantly lower theta and higher alpha activation in the audio condition. The findings provide important considerations for the design of educational VR environments.

Author Keywords
Virtual reality; Educational Technology; Learning; Cognitive Load; EEG

CCS Concepts
• Human-centered computing → Virtual reality; Empirical studies in HCI; • Applied computing → Interactive learning environments;

INTRODUCTION
Following recent advances in immersive technology, there has been a surge of hype surrounding VR in various educational contexts [21, 42, 44]. While many engaging solutions that support embodied and experiential learning are emerging [18, 32, 54, 58, 64], most immersive VR applications still include written or spoken content [13]. This is especially evident in educational systems that are aimed at teaching declarative or procedural knowledge [7, 63].

Studies in non-immersive media have highlighted that educational efficacy of a multimedia system greatly depends on the way learning content is represented to the learner. For example, empirical evidence suggests that spoken information can be more efficient than written information in certain multi-modal learning situations [40]. Similarly, it has been shown that design factors such as written and spoken information redundancy [45] or spatial and temporal distribution of learning content might influence learning outcomes in 2D media [22]. Nevertheless, little is known about if and how such findings generalize to immersive VR, which is based on an entirely different set of technological and pedagogical affordances [16, 44].

Following indications suggesting that the design of virtual reality environments influence learning outcomes [37] and that immersive technologies affect cognitive processing during learning [52, 59], we investigated the effectiveness of learning content representations in VR. Specifically, we were interested in exploring content representations for written and spoken information, which to this day dominate education [30] and remain central in commercial VR solutions (e.g., [15, 29]).

We conducted a study that combined self-reports, objective learning outcome assessments, and psychophysiological measures. Our results showed that in immersive VR representing learning content as text is superior to audio for aiding knowledge retention (i.e., acquiring and remembering basic information), but not knowledge transfer (i.e., building connections between corresponding elements of learned information). Furthermore, learners paid more attention to extraneous visual information when content was represented in audio. EEG analyses revealed significantly lower \( \theta \) and significantly higher \( \alpha \) activation for audio, suggesting reduced mental processing when learning through auditory representations in VR. Additionally, from self-reports we found that embedding textual information semantically in the virtual environment increased learners’ self-efficacy (i.e., confidence in learning) and reduced the perceived cognitive load caused by the learning material and its design.

We conclude by synthesizing our findings as design guidelines; these recommendations provide considerations for the design of learning content representations for immersive learning...
applications. Furthermore, our study demonstrates the importance of using diverse and objective measures to evaluate the effectiveness of immersive VR for learning.

BACKGROUND

We review relevant literature on learning in VR, spanning the fields of educational psychology and human-computer interaction (HCI). Specifically, we consider general trends in educational VR systems, and summarize findings regarding the effectiveness of textual and spoken content representations for learning.

Learning in VR

Educational VR is perhaps mostly known for procedural motor skill training in fields such as aviation and medicine [17, 47]. With improved accessibility and advances in visualization and interaction technologies, in the last decade we have witnessed an abundance of immersive VR applications in other educational contexts, such as safety training [8], and training of public security personnel [6]. In such learning applications, due to the unique affordances of immersive technology, VR has brought new opportunities for visualizing and interacting with abstract learning content (e.g., molecular structures [65]), as well as for simulating situations that would otherwise be dangerous to access in real life (e.g., hazardous situations) [44].

Recently, a special interest in this technology has emerged in fields reliant on declarative and conceptual knowledge acquisition (i.e., learning facts, concepts, theories, and principles). This has been particularly noticeable in learning of concepts in mathematics and physics [7, 23, 33]. Such VR applications are often reliant on written and spoken information, embedded within environments containing human-scale visualizations and animations that do not have an explicit learning value, but offer contextualization for declarative learning content (e.g., [15, 29]).

The rising interest in applying VR in education has sparked new research initiatives set out to investigate the value of educational immersive technology. Despite the increase in scholarly interest in learning in VR, some are voicing a lack of rigorous, theory-driven empirical research to shed light on how VR influence learning [12, 21, 46]. Many studies today focus on usability tests, user preferences, or reaction ratings of novel VR prototypes. While these aspects all provide relevant information about the relations between learning and immersive media, they do not inquire into the fundamental learning properties of educational VR systems.

Studies that evaluate the educational efficacy of VR, often focus on evaluating post-session learning outcomes that are reliant on participant’s self-reports. Their findings suggest that when compared to other media, VR-based instruction is more effective in various learning-relevant self-reported outcomes, such as enjoyment, self-efficacy, perceived learning and motivation [34]. Recent studies have also started to inquire into more objective measures of learning, evaluating learning outcomes such as retention of information (i.e., remember basic facts) and knowledge transfer (i.e., ability to use learned information in new contexts) [10, 34, 37]. For knowledge retention, research findings have been inconclusive. Results, however, show an emerging tendency for VR to positively influence deeper levels of knowledge comprehension [34].

Using EEG to Measure Learning

Following the quest to objectively measure learning, recent studies have also suggested the potential of using psychophysiological methods to inquire into the actual on-line cognitive processes that occur during novel information acquisition. Specifically, frequency-based analysis of electroencephalography (EEG) data have gained traction as an unobtrusive and objective measure across media, including VR [2, 37, 38, 57]. Previous empirical work in this direction has for the most part focused on measuring oscillatory activity of $\theta$ 4-8 Hz and $\alpha$ 8-12 Hz frequency bands. These frequency bands have consistently been demonstrated to be sensitive to changes in such learning-relevant cognitive processes as attention, cognitive load, and working memory [1, 4, 5]. Theta frequency activity, especially in frontal areas, has been linked to working memory capacity and to knowledge acquisition [1, 2, 51, 57]. In these studies, increasing levels of spectral power in the $\theta$ band have been proposed to reflect increasing workload. The positive relation between $\theta$ activation and mental effort has also been implicated in reviews by Klimesch [27, 28]. Increases in oscillatory $\theta$ activity, particularly in parietal sites, has also been shown to promote effective long-term memory encoding and predict later successful memory retrieval [48]. Other studies have linked higher $\theta$ activation to increased intrinsic difficulty of learning content [2, 9]. Activity in the $\alpha$ frequency band, on the other hand, is typically linked to changes in the general cognitive attention and alertness [27, 28, 53]. High levels of $\alpha$ power have been shown to reflect a state of low cortical arousal in which a decline in attention is expected. As such, $\alpha$ band power has been repeatedly observed to increase during sleep (i.e., low cortical arousal), and decrease with increased task difficulty (i.e., high cortical arousal) [2, 4, 27, 28]. In a learning context, higher $\alpha$ frequency activity has been associated with reduced cognitive load observed in gifted students [2]. In other studies that measured task performance in multitasking environments, it has been linked to increased task difficulty and cognitive demands [51]. While studies utilizing EEG for measuring learning in VR are at their infancy, initial findings have reported higher $\theta$ band and lower $\alpha$ band activation in VR learning conditions, indicating increased cognitive load in immersive environments [37]. Nevertheless, despite an acknowledged potential of using EEG frequencies for studying instructional design effects [37, 57], little is still known about how oscillatory brain activity relates to different content representation designs in educational VR.

Learning with Text and Audio

The relations between textual and auditory representations and learning have a long history of research in educational psychology, only intensified with recent advances in educational technologies. The Modality Principle, generally states that on-screen speech is superior to on-screen text for learning purposes [40]. The principle is especially evident in situations that have complex graphic representations that include dual-channel (i.e., auditory and visual) processing in working memory. According to cognitive load theory and the dual-channel processing assumption, in visually dense situations,
working memory is already overloaded by visual stimuli, and presenting some information in audio should therefore reduce cognitive load and improve learning [40, 62].

Since VR in its essence is multimodal and heavily relies on vision; when considering the modality principle, the expectation would be that learning outcomes in VR would be improved by representing information in audio. Nevertheless, since current theories are inconclusive across technology, it is unknown how previous findings from 2D media would generalize to immersive learning situations. Also, the principle has been found to be dependent on other factors (e.g., learned subject, difficulty of information, number of modalities, and division of attention [40]).

Other theories such as the Colavita visual dominance effect [60] provide alternative predictions for the effectiveness of auditory information representation. Opposing the modality principle, the Colavita effect states that the dominance of vision over audio in multisensory simulations causes visual information to be more effectively consolidated in working memory [60]. Similarly, others also argue that auditory information is inferior to written information due to its transient nature; when reading novel information, learners can reread text or adjust their reading speed to the complexity of learning content. When learning through audio, however, presented information rapidly disappears and is replaced by new information, which can result in an overwhelming working memory load [66].

**Designing Contextually for VR**

Another way to assess the effectiveness of auditory and textual learning content representations in VR is by considering how the learning content has been embedded in the VE. Virtual worlds need not necessarily to reflect the real world, and hence may incorporate content, actions, or physics that would otherwise be unrealistic. Evidence suggests that visual realism in VEs leads to increased presence (e.g., [24, 31]), and that for object manipulation in VR, high interaction fidelity is preferred [55]. Yet, whether real-world congruent VEs are superior to the implausible for learning is unknown.

Few studies have, however, explored the importance of narratorial consistent design of contexts, affordances, and interactions in virtual worlds. Sra et al. [61], for instance, found that embedding context-sensitive attractors coherent with the narrative of the VE minimized disruption and dizziness for walking in VR. Also, Harley and colleagues [20] explored the use of tangible objects with fitting narratives for VR; building on research in tangible and embodied interaction, they presented basic guidelines for tangible narratives.

A growing body of literature in embodied interaction also considers the relations between the body, the environment and cognitive activity [14, 25]. Evidence suggests that when we read, we unconsciously use spatial and haptic representations offered by the reading interface, its surrounding environment and our own physical bodies [39]. In tangible and embodied interfaces (i.e., books, print media) these factors are said to offer mental anchors to help us encode, and later decode information that we have read. In digital media, however, due to decreased materiality and bodily function, these anchors tend to disappear, thereby impeding important comprehension processes.

Such literature points towards the efficacy of realism and embodied interaction of virtual worlds, with some studies arguing for the design of context-relevant content in VR. Little actual empirical evidence, beyond the effect of presence, however, supports this view. These theories are furthermore especially undeveloped in the field of learning in VR. On a practical level, this gap in empirical evidence makes it hard to justify design decisions when incorporating learning information in educational VR systems and calls for further research to understand how the embeddedness of content might affect learning outcomes and cognition during the process of learning in VR.

**Limitations of Previous Research**

Since VR is becoming more prevalent for educational purposes, and because it is fundamentally different from other media systems (e.g., higher immersion and presence), it is important to investigate whether previous theories assessed in 2D media are applicable to the design of educational VR systems. Yet, as of today, only preliminary research has investigated the effects of instructional design in VR. Even less is known about the incorporation of text and audio in immersive VR. For instance, few studies have investigated the redundancy of textual and spoken information in VR [37, 45], and another study has investigated different UI parameters for reading in VR [13]. This lack of evidence implies that in the context of educational virtual reality, the question of how to optimally represent information is still largely unknown, both in terms of modality (i.e., text and audio) and its presentation in the VE (i.e., embeddedness in the virtual world).

Today, many educational VR studies focus on evaluating post-immersion learning with few isolated measures, failing to capture the intricate on-line processes that occur during learning [2,34]. As such, new research evaluating learning effectiveness of VR would benefit from including measures of cognitive processes that occur during learning [22, 59]. Following these directions, in this study we utilize an electroencephalogram (EEG) to measure the learners’ cognitive activity during novel information acquisition. In particular, we analyze the power spectrum of frequencies, produced by neuronal oscillations recorded during a learning activity. Even though this EEG method is not yet prevalent in educational research, there is increased focus on investigating continuous learning processes rather than relying solely on traditional post-test learning outcome measures [2, 36, 37, 57].

**EXPERIMENT**

We conducted a study to investigate how different representations of text and audio in educational VR systems might influence learning by comparing content representation modality (text/audio) and embeddedness of learning content in the virtual world (text in overlaying interface/text in a virtual book).

**Participants**

We recruited 78 university students to participate in the study (31 male; 42 female; 0 non-binary; ages $M = 25.7$ years,
$SD = 2.3$). Participants had little prior experience in the taught subject, with mean prior knowledge test scores of 1.7 ($1 = $ no prior knowledge; $7 = $ high prior knowledge). None were diagnosed with neurological or learning disabilities. Participants were recruited using an internal mailing list. The study complied to the institution’s ethics review process, hence providing participants full anonymity, and receiving their informed consent prior to the study. Participants were rewarded with a gift certificate for their participation.

**Experimental Design**
We employed a between-subjects design with one independent variable, *content representation*, with three conditions: Overlay, Book, and Auditory. Participants were randomly assigned to a condition (25/23/25 participants per condition, respectively). We chose between-subjects as the post-experiment questions would otherwise be invalid. In the Overlay condition, learning content was presented as text in an overlay interface (see Figure 1a); in the book condition the same content was presented in a virtual book (see Figure 1b); in the auditory condition, a pre recording with identical information was used (see Figure 1c).

**Research Questions**
We depart from the following three research questions.

**RQ1:** Is reading or listening superior for learning in VR? On the one hand, reading shields attention, yet is expected to be more cognitively demanding. On the other hand, auditory information is known for its transient properties.

**RQ2:** How do text interfaces in VR aid learning? Specifically, will an overlay interface compared to a book interface result in different learning outcomes?

**RQ3:** What are the differences in cognitive processing during learning between audio and text, and between different embeddings of text-based interfaces in VR?

**The VR Learning Application**
A VR learning application was developed in three different versions in Unity 2018, deployed for the HTC Vive VR system. The VR application consisted of a VE that simulated a hospital room (identical across conditions), alongside UIs with learning content.

The design of the VE was chosen to contextualize the learning content. The VE was populated with two humanoid avatars (a doctor and a patient) and objects relevant to a hospital room (i.e., hospital bed, painting, TV, clock, etc.). The doctor character had idle animations applied. Ambient hospital sounds were present in the background. Participants were not embodied by an avatar and could not interact with any of the contextualizing objects in the VE.

Sarcoma cancer was chosen as the learning topic of the application, due to the general public’s relative unfamiliarity with the subject. We adapted the content from a pamphlet developed by the National Cancer Society, targeted the general public.

Information was delivered in participants’ native language (Danish) through 25 paragraphs of approximately equal lengths (300-400 characters), one at a time. For the auditory condition, a human voice actor recorded the information (at 146 words per minute on average). To switch to the next paragraph (both text and audio), users pressed the trigger button on a handheld controller. To control for learning content exposure, participants were not allowed to go backwards.

In the two textual conditions (Overlay and Book), text-formatting features (e.g., font family and size, hyphenation) were identical. For the Overlay condition, a semi-transparent canvas was positioned in a static position in the VE (see Figure 1).
Measurements

The dependent variables (see Table 1) were acquired in a survey administered immediately after the VR experience on a desktop PC; EEG data was collected throughout the entire simulation. A pre-session survey was administered before the learning experience to gather information on demographics, prior knowledge, education, and previous VR experience.

Learning outcome measures

Following guidelines from educational psychology (e.g., [34, 43, 45]), two custom-made tests were developed for evaluating learning outcomes—a knowledge retention test and a knowledge transfer test. In the retention test, 25 multiple choice questions were designed to assess how much of the experience information could be remembered, measuring basic levels of knowledge acquisition. In the transfer test, three open-ended questions were employed, where participants had to apply the acquired knowledge to a novel situation to assess deeper levels of knowledge comprehension. Two independent coders graded the transfer tests (1-4 points per answer); conflicts were resolved through discussion. To measure participants’ self-reported sense of confidence in the learning experience to gather information on demographics, prior knowledge, education, and previous VR experience.

Self-reported cognitive load measures

To measure self-reported cognitive load we employed four measures: a mental effort item from Paas [49], aimed at assessing cognitive capacity invested during learning; an item to measure perceived concentration during learning, adapted from Salomon [56]; an intrinsic cognitive load item from Ayres [3] to measure the perceived difficulty of the subject; an extraneous cognitive load item from Cierniak et al. [11] to assess how difficult the participants experienced the instructions. These measures were all single items, administered on a 9-point Likert scale.

Visual attention measure

To understand cognitive resource allocation further, we developed an extraneous attention measurement, where five open-ended questions were used to evaluate how much attention was paid to irrelevant elements in the VE, peripheral to learning (e.g., “There was a painting hanging across from you in the hospital room - which object was drawn on the painting?”). Answers were scored following the same procedures as the transfer test.

Psychophysiological measures

EEG data was collected using Advanced Brain Monitoring (ABM) X-10, wireless 9-channel system running at 256 Hz. Nine Ag/AgCl EEG electrodes were located at Fz, F3, F4, Cz, C3, C4, POz, P3, P4, according to the international 10-20 system [26], referenced to linked mastoids, with impedance levels below 10Ω. To isolate cognitive activity occurring only during information acquisition phases, EEG data collection was triggered by a button press on the HTC Vive controller. To ensure synchronization between stimulus representation, triggering and EEG data, an ABM External Sync Unit (ESU) and a Cedrus StimTracker were employed. EEG data was recorded using iMotions biometric data acquisition software. EEGlab toolbox for MATLAB was used to filter the gathered EEG data (0.5 Hz high-pass, 100 Hz lowpass, CleanLine at 50 Hz and 100 Hz) and to perform MARA-guided independent component analysis (ICA). Building on learning studies using EEG [2, 28, 37, 57], we focus on the θ (4-7 Hz) and α (8-12 Hz) frequency bands, which have been previously related to cognitive load during learning.

Procedure

Participants were given a walk-through of the protocol before participating, and were asked to take notice of the environment during learning. The study proceeded as follows: (1) participant briefing and consent signing, (2) EEG mounting, (3) standardized EEG signal quality and impedance testing, (4) pre-session survey, (5) introduction to VR experience and controls, and HMD mounting, (6) VR immersion (approx. 20 minutes), (7) removal of the EEG and HMD, (8) post-session...
survey, and (9) participant debriefing. During the VR experience, participants were seated in a comfortable position and were asked to minimize unnecessary body/facial movements to lessen the possible signal artifacts in the EEG recordings. Each session took about 75 minutes.

RESULTS
Using a $\chi^2$ test, we found no significant differences between the three conditions attributable to age ($p = .82$), gender ($p = .15$), educational level ($p = .65$), or VR experience ($p = .99$), measured in the pre-session survey. Next, we report the findings on the dependent variables, using descriptive statistics and relevant statistical testing.

Learning outcomes
We employed three learning related outcome measures. We applied parametric analyses using one-way ANOVAs and subsequent post-hoc Tukey’s HSD tests, where applicable. Figure 2 shows mean values with confidence intervals.

We found significant differences for retention and self-efficacy attributable to the experimental condition, however not for transfer. For retention we found a significant difference, $F(2,70) = 4.7$, $p = .01$, with the post-hoc test showing that Overlay-Audio ($p = .04$) and Book-Audio ($p = .02$) comparisons significantly differed. For self-efficacy we also found a significant difference: $F(2,70) = 4.8$, $p = .01$, the post-hoc test showed a significance for the Book-Audio ($p = .01$) comparison, and borderline for the Overlay-Book ($p = .05$) comparison.

The results show that reading compared to listening in VR was superior in terms of retention (i.e., remembering learned information), yet we did not find an effect for transfer (i.e., using learned information in a new context). The study showed that self-efficacy (i.e., confidence in one’s learning) was reported highest in the book condition.

Self-reported Cognitive load
The four cognitive load measures were analyzed using a non-parametric Kruskal-Wallis rank sum test, as these were single question items. Where applicable, Bonferroni-corrected post-hoc Dunn’s tests were used to test which groups significantly differed. Figure 4 therefore shows medians and the interquartile ranges (IQR).

No significant differences were found for mental effort or concentration. However, intrinsic cognitive load showed to significantly vary with the experimental condition, $\chi^2 = 7.85$, $p = .02$; the post-hoc test showed that intrinsic load was reported lowest for the book condition, with a significant difference found between Book and Audio ($p = .01$). Book also showed lower intrinsic load than Overlay, yet this difference was not significant ($p = .06$). For extraneous cognitive load, the difference was also significant, $\chi^2 = 7.22$, $p = .03$; the post-hoc test showed that Book and Overlay were significantly different ($p = .01$).

The results show that differences in cognitive load are attributable to the learning content representation. It is evident that the Book condition yielded the lowest intrinsic and extraneous cognitive load, meaning that the contextually embedded VR design reduced the perceived difficulty both due to the subject and the instruction format.

Attention
Most extraneous attention was paid in the auditory condition. Extraneous attention showed to significantly vary between conditions: $F(2,70) = 4.1$, $p = .02$, with the post-hoc test showing that Overlay and Audio significantly differed ($p = .03$), while the difference between Book and Audio was not significant ($p = .06$). This shows that participants who received the taught information in audio compared to text in Overlay, payed more attention to extraneous objects in the VE.

EEG measures
Power Spectral Density (PSD) estimates were constructed from continuous EEG data for each electrode for all participants, using the Discrete Fourier Transform (DFT), with a Hanning window of 1 sec width and 50% overlap. EEG data was normalized and log-transformed to minimize skewness.

Figure 2: Three measures of learning outcomes. The figure shows mean values, with bars showing bootstrapped 95% confidence intervals.

Figure 3: Mean extraneous attention, with bars showing bootstrapped 95% confidence intervals.
and to standardize unit variance. Analysis was facilitated by using the Neurospec toolbox for Matlab [19]. Mean peak frequency estimates were further calculated per frequency band (4-7 Hz for $\theta$ and 8-13 Hz for $\alpha$) and compared between the three experimental conditions using one-way ANOVAs.

For the mean $\theta$ EEG values we find significant differences between conditions in Fz, Cz and parietal electrodes POz, P3, and P4; see Figure 5 left; $p = [1^{-0.04}]$. We recorded significantly lower $\theta$ activity for the audio condition for all of five electrodes, indicating that listening compared to reading in VR yields lower levels of memory workload. No significant differences were found between the two text conditions.

For the mean $\alpha$ frequencies we find significant differences between conditions on seven out of nine electrodes (F3, Cz, C3, C4, POz, P3, and P4; see Figure 5 right; $p = [0.001;0.05]$). No differences were found between the two text conditions.

We recorded significantly higher $\alpha$ activity for the audio condition for all of seven electrodes, indicating an overall lesser cognitive effort when learning through auditory content.

**DISCUSSION**

In the current study we set out to explore the effectiveness of textual and auditory information displays for learning in immersive VR. We first discuss findings that compare visual and auditory learning information representations, and then review our results related to embeddedness of educational content in a VE.

**Textual and auditory representations in VR**

The major finding of this study is that textual representations in VR are superior to auditory representations in terms of knowledge retention, but no differences between conditions were found for knowledge transfer. Prior literature has related retention tests to the construction of fundamental representational connections (i.e., acquiring and remembering presented information), while transfer tests are known to involve both representational and referential connections (i.e., building connections between corresponding elements of presented information, problem-solving) [41]. This means that even though participants were able to form general references for building an overall understanding from the learned material, they had
difficulty in acquiring basic information, and in remembering specific factual details when learning with audio.

One explanation for the learning outcome results could be extrapolated from the EEG findings, where an overall significantly higher $\alpha$ activity and significantly lower parietal $\theta$ activity (P3, P4, POZ) were observed in the audio condition. Since $\alpha$ suppression is known to relate to general cognitive attention [28], this result suggests that participants invested overall less cognitive effort when learning with audio. Such lack of cognitive engagement might have led them to acquire fewer factual details, as reflected in the retention test results. Nevertheless, increases in parietal $\theta$ have been previously associated with long-term memory encoding and retrieval [48]. In our study, participants in the auditory condition were able to build comparable referential connections needed to answer the transfer test, but these EEG results also suggest that the participants had less long-term memory processing during learning. An explanation could be that even though the transfer test used in this study asked the participants to make references between the different learned concepts, it did not succeed in isolating long-term memory retrieval processes, which were captured by the EEG. This highlights the importance of using multiple methods when inquiring into cognitive processes during complex tasks, such as learning in immersive technologies [34, 35].

It is also worth noting that previously literature has linked higher levels of $\theta$ band power to increased intrinsic difficulty [2, 9]. In this study, however, even though participants had lower $\theta$ activity in the auditory condition, they reported perceiving the learning content to be more difficult when learning with audio (i.e., intrinsic cognitive load). As the intrinsic difficulty of the learning content did not actually change across conditions, the perceived increase in self-reported difficulty in the auditory condition could be reflected by the transient nature of this format [66]. In the auditory condition the participants might not have been able to actively engage in working memory processes (i.e., selecting and organizing essential elements of the content) to the same degree as with the static textual representations, where the participants were able to more easily repeat and integrate information across sentences.

Another interesting finding comes from the results for extraneous visual attention, where participants in the auditory condition remembered significantly more details about the peripheral objects in the VE than in the Overlay condition. Even though no learning-relevant information was embedded in those objects in this study, results show that participants invested cognitive resources in acquiring information from the VE and successfully remembered it in the extraneous attention test. Considering that EEG frequency analyses showed an overall lower cognitive effort for the audio condition, these results support the dual-channel processing assumption [40], suggesting that in visually dense conditions (i.e., immersive VR) there is a tendency to split working memory capacity between the visual and auditory channels. Nevertheless, given that learning outcomes were lower in the audio condition, and that participants paid more attention to visually presented information (i.e., written text and peripheral visual information in the VE) our findings draw more parallels to the Colavita visual dominance effect and less to the Modality principle. This is additionally supported by the results gained for self-efficacy and intrinsic cognitive load, where the participants reported lower confidence in their learning and perceived learning information as more difficult in the auditory condition.

**Embedded and non-embedded text representations in VR**

Regarding the embeddedness of textual information in a VE, important findings were observed for self-efficacy and self-reported cognitive load. For self-efficacy, participants felt significantly more confident about their learning when they learned from the contextually embedded interface (i.e., a virtual book), than from the contextually non-embedded interface (i.e., overlay). Given that self-efficacy is known to be one of the main precursors for educational success [50], our results suggest that learning content embeddedness might be an important factor to consider when aiming to provide psychological and metacognitive support for learners.

Another important finding comes from our results for self-reported extraneous cognitive load, where participants reported that it was more difficult to read from an overlay interface than from a virtual book. Factors such as the learner’s proximity to the reading interface might have influenced this result [13]. However, since no differences between these conditions were observed in any of the objective measures, we find this explanation implausible. One explanation could be that since VR is considered to be an embodied medium [54], participants might find it easier to read from a more embodied reading interface that resembles the affordances of real-world reading. The efficacy of embodiment for reading has previously been considered in non-immersive media, where findings suggest the superiority of embodied reading interfaces (e.g., books) compared to non-embodied interfaces (e.g., computer screens) [39].

**Limitations and Future Work**

The scope of this study was focused on providing empirical evidence for the effectiveness of different learning content representations, and less so on investigating the underlying theoretical assumptions behind the presented theories (i.e., Colavita visual dominance effect; the Modality principle, and Embodied Cognition theories). Additional considerations would be needed to draw conclusions in this regard. As an example, future research could for example consider integrating some of the learning information into the surrounding VE, as splitting learning information between different representation modalities might be important for achieving the modality effect [40]. At the same time, since this study did not specifically inquire into the actual affordances of embedded text versus non-embedded text for reading in VR, future research should consider more in-depth investigations in this direction. Furthermore, the current study employed an extraneous attention measure that did not provide comprehensive insight into the actual on-line attention fluctuation, and instead focused on measuring recall of extraneous objects that could indicate extraneous attention. To attain a more objective measure of on-line processes of attention, future studies could therefore benefit from the use of eye-tracking measurements [2, 34].
addition, this instrument only measured extraneous information that was presented visually, but considering extraneous information also in the auditory modality could yield a more comprehensive understanding of sensory dominance [60].

Another limitation of this study deals with the fact that the participants were asked to limit their bodily movements during their learning experience in VR. While this benefited psychophysiological data collection, it is important to note that it might also have influenced their overall interaction patterns in the environment.

Besides this, in the audio condition the participants could not rewind and re-listen to auditory information, while in the text-based conditions could re-read on a per-page basis. While this was necessary for triggering procedures of EEG data, and while we expect that no participants read the text more than once, the study could be improved by utilizing eye-tracking that would allow us to control for re-reads of text. Another way to improve this aspect would be to pre-define and limit the amount of time the participants could interact with text interfaces. However, this would compromise the ecological validity of this study.

Finally, it could be argued that the chosen topic of Sarcoma cancer and utilized information representations (i.e., textual and audio) do not particularly leverage on all of the affordances of immersive VR. Nevertheless, these choices were consciously made in this study with the aim to highlight that there is an ongoing trend in educational technology to remediate content from non-immersive to immersive media, without taking particular care of how this content will be experienced in its new form [38].

DESIGN CONSIDERATIONS
Besides providing empirical foundation for the importance of modality and context in educational immersive technology, our study revealed a range of findings that researchers and practitioners could consider when designing VR learning applications. Here we attempt at summarizing the most important findings to provide applicable design recommendations (see Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Text</th>
<th>Audio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Remember facts</strong></td>
<td>Higher</td>
<td>Lower</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Deep knowledge</strong></td>
<td>No difference</td>
<td></td>
<td>.2</td>
</tr>
<tr>
<td><strong>Self-efficacy</strong></td>
<td>Higher</td>
<td>Lower</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Difficulty of subject</strong></td>
<td>Lower</td>
<td>Higher</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Difficulty of instructions</strong></td>
<td>Lower</td>
<td>Higher</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Attention elsewhere</strong></td>
<td>Lower</td>
<td>Higher</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Memory workload, θ</strong></td>
<td>Higher</td>
<td>Lower</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Cognitive activity, α</strong></td>
<td>Higher</td>
<td>Lower</td>
<td>.05</td>
</tr>
</tbody>
</table>

Table 2: Findings summarized: dependent variables and their differences across three VR learning designs, spanning two modalities (text/audio) and two visual interfaces (overlay/book).

Text or Audio in VR
For learning in VR, we provide empirical evidence suggesting the superiority of textual representation of factual information, in terms of number of facts remembered (i.e., retention). Our results suggest that text compared to audio in VR makes information more available to the learner, with lower reported difficulty scores for extraneous and intrinsic cognitive load. However, there might still be reasons for incorporating instructional audio in VR-based learning applications. Specifically, extraneous attention scores were consistently higher for the audio condition, showing that participants were able to recall visual information from the VR more successfully than in the other two conditions. Similarly, EEG analyses showed that audio yields significantly lower mental memory load and effort, suggesting that audio might not be as cognitively demanding as textual information.

Embedding Text in VR
In our study, we compared reading from an overlay interface and through a virtual book. While these two UIs did not differ in knowledge acquisition, the participants who read the material through the book reported lower cognitive load and higher self-efficacy, which for most learning scenarios would be an advantageous ideal.

Considerations
Some applications might have deliberate unconventional design goals, yet for VR learning applications following the same principles as the study presented in this paper, the following three design considerations could be taken into account:

1. Important factual learning information should be presented as text
2. Text should be contextually embedded in the environment
3. Learning information could be presented in audio if attention to the environment is required, or if memory workload should be minimized
4. Learning information could be presented in audio if remembering facts is not a goal for the simulation

CONCLUSION
In this paper we presented a between-subjects study investigating the learning outcomes and processes across designs in an educational VR application. The results of the study indicate that different ways of representing information in VR can influence learning. With objective learning measures, self-reports, and EEG we report differences across attending knowledge from text in an overlay interface, in a virtual book, and from audio. The findings are then synthesized to design considerations for educational VR tools. The findings are especially relevant for the design of immersive educational VR systems.

Furthermore, our study may pave the way for future empirical investigations across HCI, educational psychology, and instructional design, of how design parameters influence learning with immersive technologies.
REFERENCES


