



# What variables are connected with system usability and satisfaction? Results from an educational virtual reality field trip

Maximilian C. Fink<sup>\*</sup>, Volker Eisenlauer, Bernhard Ertl

Universität der Bundeswehr München, Department of Human Sciences, Learning and Teaching with Media, Neubiberg, Germany

## ARTICLE INFO

**Keywords:**  
Virtual reality  
Usability  
Satisfaction  
Cognitive load

## ABSTRACT

Exploring the usability of educational virtual reality (VR) is essential. Research in this area can explain the adoption of VR as a new technology, contribute to examining paths toward effective learning, and provide recommendations for effective educational design. This study investigated usability in a sample of  $N = 64$  university students learning about simplified construction engineering topics from an interactive VR field trip. Two research objectives were pursued. First, the level of achieved usability was examined using a mixed-methods approach, including data from semi-structured interviews and a survey on system usability and satisfaction. Second, the relationships between several connected variables (ease of use, usefulness, presence, and cognitive load) and system usability and satisfaction were investigated quantitatively. With respect to the first research objective, system usability and satisfaction were evaluated positively in the interviews and rated highly in the surveys. These results indicate that VR field trips, which are increasingly replacing excursions, are regarded as an appropriate instructional method. Concerning the second research objective, ease of use, usefulness, presence, and extraneous load were found to predict system usability. Further, the factors of usefulness, presence, and intrinsic cognitive load helped explain the variance in satisfaction. These findings highlight that the variables connected to system usability and satisfaction in VR learning vary. In addition to these and other theoretical implications, practical implications and recommendations for educational design are discussed.

## 1. Introduction

Virtual reality (VR) technology is rapidly gaining in popularity. Reports estimate that the number of VR devices sold worldwide will grow from 25 million in 2020 to 130 million in 2025 (Statista, 2022). Although VR possesses vast potential for learning and instruction, it still lags in terms of adoption and continued use for educational purposes (Radianti et al., 2020). One key reason for this logjam may be VR's limited usability and satisfaction (Kavanagh et al., 2017). This paper sheds light on the relationships between usability and several connected variables. A deeper understanding of this topic will help increase the adoption of VR applications in education, uncover pathways toward effective learning, and provide recommendations for effective educational design.

### 1.1. Virtual reality and its use in education

VR involves various technologies enabling users to delve fully into a virtual environment. The most popular and convenient class of

technological device that delivers VR is the head-mounted display (HMD). HMDs are goggles with high-resolution displays that shield the user from distracting stimuli of the outside world and typically track head and hand movements to represent the user in a virtual space. HMD-based VR learning differs from learning with desktop computers, because HMDs provide a multi-sensory experience including visual, audio, and haptic stimuli, allow natural and swift interaction with the virtual world, and convey a strong feeling of immersion (Mikropoulos & Bellou, 2006). VR is frequently used in fields such as engineering education, health-related education, science education, and general education (Kavanagh et al., 2017; Radianti et al., 2020). The goals pursued with VR in these fields include conveying knowledge, practicing skills, and solving problems, as well as creating positive emotional experiences (Jensen & Konradsen, 2018; Radianti et al., 2020). In the following section, we define usability and explain its significance for educational VR.

<sup>\*</sup> Corresponding author.

E-mail address: [Maximilian.Fink@unibw.de](mailto:Maximilian.Fink@unibw.de) (M.C. Fink).

<https://doi.org/10.1016/j.cexr.2023.100043>

Received 17 July 2023; Received in revised form 6 October 2023; Accepted 10 October 2023

Available online 30 October 2023

2949-6780/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1.2. Usability and its relevance in educational VR

According to a prominent, relatively technical definition provided by the International Organization for Standardization (ISO, 2018), usability is the degree to which a system—such as a VR learning environment—can be used efficiently, effectively, and conveniently for the participants' intended objectives. This construct can be operationalized with various indicators, including system usability and satisfaction (ISO, 2018). Before describing these two indicators in the next subsection, we briefly discuss why usability is important for educational VR.

The relevance of usability in educational VR learning environments becomes apparent from three points of view. From a cognitive perspective, usability can influence instructional effectiveness. VR learning environments must possess a certain level of usability so that users can carry out learning processes efficiently. A lack in usability can directly distract from learning or necessitate spending resources for engaging with the system itself (Ardito et al., 2006). From an affective standpoint, usability can influence emotions and moods. Studies indicate that usability is related to motivation, enjoyment, and presence, which are, in turn, associated with perceived learning and satisfaction using VR (Lee et al., 2010; Makransky & Petersen, 2019). From a behavioral angle, usability is critical to adopting new educational technologies. With greater usability, users may be more inclined to engage with and continue to use new educational technologies, for instance, in self-regulated learning (Nagy, 2018).

## 1.3. System usability and satisfaction

*System usability* is a global subjective rating of the usability of a system; it comprises aspects such as the level of necessary support and system complexity (Brooke, 1996). The *System Usability Scale* (Brooke, 1996) is frequently used to assess this variable. Scores ranging from 0 (*very low usability*) to 100 (*perfect usability*) quantify the system usability. Scores above 70 are rated as acceptable and point to suitability for long-term use (Brooke, 1996). A recent literature review on the System Usability Scale highlights that it has been primarily used in studies on learning management systems and learning with smartphones, but rarely in studies with HMDs (Vlachogianni & Tselios, 2022). Decent usability with a score above 70 has been reported for learning with mobile applications and learning with multimedia (Vlachogianni & Tselios, 2022). Two educational studies on VR using HMDs have examined complex learning environments using the System Usability Scale. In one of these, Huang et al. (2021) observed marginal system usability ( $M = 64.00$ ) of a VR application for training 3D modeling of products. In the other, Othman et al. (2022) discovered acceptable system usability ( $M = 72.10$ ) for users exploring a cultural heritage museum with HMDs. These studies' results suggest that usability varies in educational VR applications and may depend on a range of factors, such as the system's didactical design and content.

*Satisfaction* refers to various aspects that emerge from the congruence among user expectations, usage requirements, and user experience (ISO, 2018). According to the ISO definition, these aspects include physical responses (e.g., motor behaviors induced by a system), emotional responses (e.g., moods evoked by a system), and cognitive responses (e.g., certain attitudes towards a system). A self-report scale developed by Chou and Liu (2005) is often used to measure satisfaction in educational VR research. This scale evaluates emotional and cognitive responses during learning. Concerning empirical results, the following outcomes can be reported. Several studies exploring educational VR within diverse contexts, such as health sciences, science education, and museum education with various instructional approaches, including inquiry learning and observational learning, indicate that high levels of satisfaction can be reached (Chang et al., 2022; Chao et al., 2023; Makransky et al., 2019; Yang et al., 2022). These results align with domain-specific findings relevant to this study. Dos Anjos et al. (2021) examined through a literature review how satisfied engineering students

were with VR applications (Dos Anjos et al., 2021). Ten out of the twelve studies included in this review reported that university students were satisfied with using VR and preferred it over traditional teaching methods (Dos Anjos et al., 2021). Next, we discuss the broader topic of technology acceptance theories.

## 1.4. Technology acceptance theories

The technology acceptance literature has investigated for more than three decades how new educational tools and devices are adopted and used (Marangunić & Granić, 2015). Within this literature, the *Technology Acceptance Model* plays a crucial role in its different versions: TAM1, TAM2, and TAM3 (Davis, 1989; Venkatesh & Bala, 2008; Venkatesh & Davis, 2000). TAM1 (Davis, 1986, 1989) was developed mainly based on the postulation that the variables perceived usefulness and perceived ease of use affect usage intentions and usage behavior. *Ease of use* is the degree to which interaction with a system is perceived as effortless (Davis, 1989). *Usefulness* refers to the extent to which a system is rated as valuable or helpful for a critical objective, such as learning (Davis, 1989). TAM2 (Venkatesh & Davis, 2000) complemented TAM1 with a set of new variables devoted to social influences (e.g., subjective norm) and instrumental processes (e.g., job relevancy). TAM3 (Venkatesh & Bala, 2008) went beyond these models by incorporating user experience, belief-related variables (e.g., computer anxiety), and other individual variables (e.g., perceived enjoyment). The different versions of the TAM detail the relationships among the mentioned variables and have meaningfully explained why formerly new technologies such as tablets and smartphones gained widespread use (Alsharida et al., 2021).

Another relevant model in the literature is that of Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003). This model integrates key suppositions and findings from eight different acceptance frameworks. According to this model, behavioral intentions and usage behavior are mainly determined by performance expectancy, effort expectancy, social influences, facilitating conditions, gender, age, experience, and voluntariness of use. Some of the variables contained in the TAM are in the UTAUT model part of other variables. Perceived usefulness is considered a subconstruct of performance expectancy, whereas perceived ease of use is included under effort expectancy. An evaluation study using different types of software used in the workplace has demonstrated that the UTAUT predicted intention to use the software better than other frameworks, including TAM and TAM2 (Venkatesh et al., 2003).

## 1.5. New approaches to measure usability and empirical evidence on user acceptance

Since TAM and UTAUT were published, further advances have been made in measuring and predicting user acceptance in educational VR. For instance, Ustun et al. (2023) published an instrument based on the UTAUT to assess user acceptance in VR. This instrument has four subscales: performance expectancy, social influence, effort expectancy, and facilitating conditions. The instrument displayed good psychometric properties in two evaluation studies with college students who had experience in using VR. Further, Karaoglan-Yilmaz et al. (2023) introduced an attitude scale for educational VR settings. This scale contains nine items and mainly focuses on learners' positive attitudes. It can be used in evaluation studies and scientific investigations as a baseline measurement of attitudes or for pre-post comparisons of attitudes. The instrument was successfully evaluated for university students and displayed high reliability.

Considerable empirical evidence is available on user acceptance. A meta-analysis of 42 studies from the e-learning literature reveals that ease of use and usefulness possess, on average, medium relations to different usability facets (Šumak et al., 2011). More specifically, positive correlations were found between ease of use, usefulness, attitudes towards using, and behavioral intentions (Šumak et al., 2011). In addition

to these findings, a study has shown that ease of use and usefulness are linked with satisfaction in mobile learning (Ohk et al., 2015). Despite this wealth of empirical findings, not all relationships are yet clear. A recent literature review shows that empirical findings on the system usability scale and variables contained in the technology-acceptance literature are missing, particularly for VR with HMDs (Vlachogianni & Tselios, 2022).

### 1.6. Presence and cognitive load

Frameworks on learning within VR (Makrasky & Petersen, 2021; Vogt, 2021) posit that the variables *presence* and *cognitive load* are also related to usability in VR. We define these variables in this section and explain their connection with usability.

*Presence* is “the subjective experience of being in one place or environment, even when one is physically situated in another” (Witmer & Singer, 1998, p. 225). Studies on immersive and desktop-based VR learning show that presence is associated with increased interest and motivation (Makrasky & Lilleholt, 2018; Makrasky & Petersen, 2019). Moreover, several studies have demonstrated that presence and usability—measured with varying instruments—are linked. Voinescu et al. (2023) and Wienrich et al. (2018) found support for this link for VR with HMDs. Brade et al. (2017) report evidence for this link in a context where VR was used with a CAVE system containing multiple displays surrounding the participants. In addition, a theoretical perspective proposes that increased flow experiences and similar states like presence can positively impact user satisfaction. In line with this reasoning, a study in which participants learned from a desktop-based VR application discovered that higher presence perception was linked positively with higher satisfaction (Liu et al., 2023).

*Cognitive load* entails the demands placed on working memory when learning. According to Sweller et al. (1998), three types of cognitive load must be distinguished: *Intrinsic load* is mainly determined by the amount of material to be learned and learners’ prior knowledge. *Extraneous load* results from the way the content is conveyed in the learning environment and is mainly influenced by instructional design. *Germane cognitive load* emerges from processes of schema construction during learning. It has been established that cognitive load is particularly high in HMD-based VR compared to desktop-based VR (Makrasky et al., 2019) and that higher extraneous cognitive load is associated with reduced knowledge acquisition in VR with HMDs (Albus et al., 2021). Based on Kalyuga’s (2011) specifications for intrinsic and extraneous load, the following can be assumed. When the difficulty of the material is considered too high or too low (inadequate intrinsic load), participants may become frustrated or bored, and their satisfaction might decrease. When the learning environment is perceived as overly complex (heightened extraneous load), search processes may be required, and interactions with the system could become intricate (Kalyuga, 2011). Therefore, cognitive load and usability are related concepts. In fact, Hollender et al. (2010) operationalize extraneous load to stem from educational design choices and users’ interaction patterns. Thus, learning environments with distracting and unnecessary information and complicated interaction patterns may reduce usability (Hollender et al., 2010).

### 1.7. Research gap, research questions, and hypotheses

The usability of VR learning environments has been researched extensively under two largely separate research traditions. The computer science subfield of human–computer interaction primarily includes studies on system features and their interaction with experienced usability. The field of educational psychology, however, has been primarily interested in cognitive processes during learning and usage intentions. We are convinced that through integrating both research traditions, we can obtain many valuable findings to close the research gap in this area. To make a valuable contribution to the relevant

literature, we investigate the interaction of cognitive load and usability and examine the link between presence and usability in a VR learning environment.

Another critical aspect of the research gap is the lack of mixed-methods approaches. The large majority of studies we reviewed used surveys that were analyzed quantitatively. Mixed-methods research makes additional insights accessible through combining quantitative and qualitative analyses. This type of research provides opportunities to test known theories but is also helpful for idea generation and theory development (Almalki, 2016).

Lastly, rapid technological progress requires us to gain up-to-date insights into the usability of VR learning environments. Only a few years ago, the usability of VR learning environments was often poor due to hardware and software issues (Kavanagh et al., 2017). Now that HMDs and VR software have undergone many improvements, the question is whether adequate usability can be achieved in educational applications today. Against this theoretical background, and considering the mentioned research gaps, we investigate four research questions in this study.

Research questions RQ1 and RQ2 address the usability and satisfaction attained in a VR learning environment. These research questions are investigated using a mixed-methods approach. First, data from semi-structured interviews are analyzed qualitatively. Then, the results are reexamined through quantitative analyses of survey data.

**RQ1:** *What level of system usability does the VR learning environment reach?*

**H1.** More than an acceptable level of system usability (score >70, the cut-off for an acceptable score) is reached.

This hypothesis assumes that decent usability scores reported for various educational technologies by Vlachogianni and Tselios (2022) can also be reached in educational VR with HMDs. It aligns with results from a VR study by Othman et al. (2022) and is partially in contrast to lower system usability discovered in a VR study by Huang et al. (2021).

**RQ2:** *What level of satisfaction does the VR learning environment reach?*

**H2.** More than an acceptable level of satisfaction (score >3.0, the theoretical scale mean) is reached.

Empirical support for acceptable levels of satisfaction comes from numerous VR studies in engineering education (Dos Anjos et al., 2021) and VR studies from other domains using different instructional approaches (Chang et al., 2022; Chao et al., 2023; Makrasky et al., 2019; Yang et al., 2022).

Research questions RQ3 and RQ4 contribute to unraveling how connected variables are linked with system usability and satisfaction. These research questions are examined quantitatively based on survey data.

**RQ3:** *To what extent do connected variables explain system usability in the VR learning environment?*

**H3.1–H3.4.** Ease of use, usefulness, presence, and extraneous load explain system usability.

Support for the hypotheses on ease of use and usefulness can be found in meta-analyses and reviews on technology acceptance and system usability (Vlachogianni & Tselios, 2022; Sumak et al., 2011). For the assumed relations with presence and cognitive load, evidence can be found in empirical studies and theoretical contributions (Brade et al., 2017; Hollender et al., 2010; Kalyuga, 2011; Voinescu et al., 2023; Wienrich et al., 2018).

**RQ4:** *To what extent do connected variables explain satisfaction in the VR learning environment?*

**H4.1–H4.4.** Ease of use, usefulness, presence, and intrinsic load explain satisfaction. The hypotheses on ease of use and usefulness are supported by a review on technology acceptance (Sumak et al., 2011) and a study by Ohk et al. (2015). The hypotheses on presence and



intrinsic load are based on results from a primary study by [Liu et al. \(2023\)](#) and theoretical arguments ([Kalyuga, 2011](#); [Witmer & Singer, 1998](#)).

## 2. Method

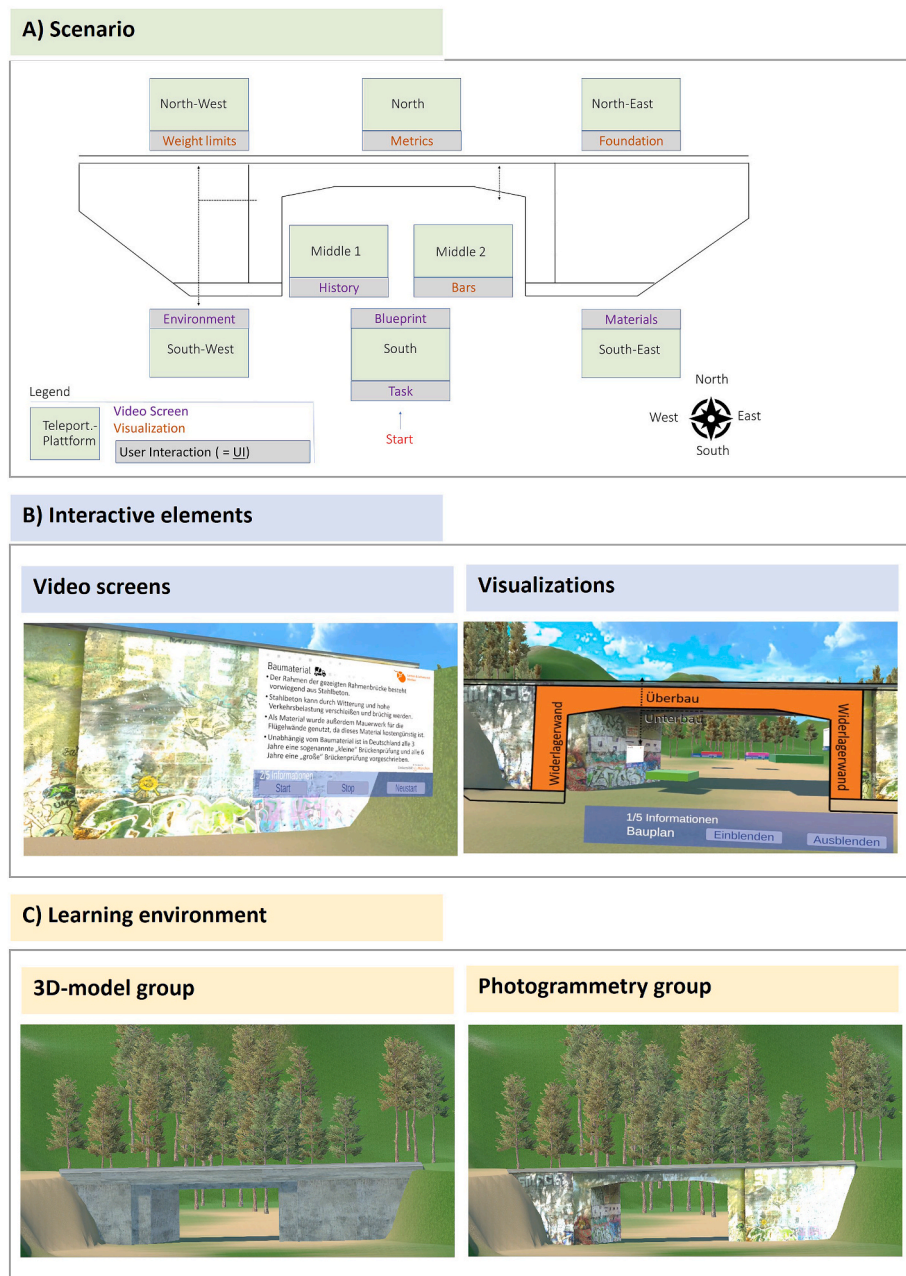
### 2.1. Participants

The study survey was completed by  $N = 64$  educational sciences and psychology students. Due to data protection regulations, age was assessed in broad categories, and the distribution is as follows: 17.2% ( $n = 11$ ) below 21 years, 59.4% ( $n = 38$ ) between 21 and 24 years, and 20.3% ( $n = 13$ ) above 24 years, with data missing for 3.1% ( $n = 2$ ). Most participants had very little (81.2%,  $n = 52$ ) or little (14.1%,  $n = 9$ ) VR experience; a few participants skipped this question (4.7%,  $n = 3$ ). In

addition, a subsample of  $n = 8$  participants completed a semi-structured interview. Participation in the study was voluntary, and participants received a compensation of €5 or extra credit in a class for taking part.

### 2.2. Learning environment

For the study, we selected the domain of construction engineering, because VR can be a powerful learning aid in this domain. In fact, several studies on engineering education have shown that VR can bolster students' motivation and comprehension through visualizing abstract blueprints and complex building structures ([Lanzo et al., 2020](#); [Wang et al., 2018](#)). The learning environment was based on the didactical format of a VR field trip, which can be defined as “a journey taken without actually making a trip to the site” ([Woerner, 1999](#), p. 5). It conveyed basic knowledge about bridges and their characteristics, and



**Fig. 1.** The main features of the VR field trip were adapted from [Fink et al. \(2023a\)](#)

The scenario comprised seven teleportation platforms about bridges and their characteristics. Users controlled the learning environment with interactive elements, and the study was conducted with a 3D model group and a photogrammetry group.

the VR content was adapted for university students of non-technical subjects. With respect to the selected topics, the learning environment included, for instance, construction plans and building materials.

Participants had the task of exploring a railway bridge in the learning environment and collecting information about it. The scenario was as follows (see Fig. 1 [A]): Participants started at the first teleportation platform and then navigated freely to seven other platforms. Each platform had interactive elements (see Fig. 1 [B]). Participants could either display a visualization such as a figure of a blueprint that enhanced their understanding or watch a video with a presentation about a construction engineering topic.

### 2.3. Instruments

This section describes the instruments used to assess the different usability facets and their connected variables. These instruments were adapted for use in a VR study.

*Ease of use* and *usefulness* were surveyed with two scales developed by Lee et al. (2010). Both scales consist of four items measured on a Likert scale with scores ranging from 1 (*not agree at all*) to 5 (*fully agree*). Ease of use was assessed with items such as “Learning to operate this type of virtual reality learning environment is easy for me.” Usefulness was investigated with items such as “This type of virtual reality learning environment is useful in supporting my learning.” Total scores were evaluated by calculating the scale mean. The internal consistency was determined as  $\alpha = 0.67$  for ease of use and  $\alpha = 0.83$  for usefulness.

*Presence* data were gathered using the *Igroup Presence Questionnaire* (Schubert et al., 2001). This questionnaire includes 13 items that focus on the involvement, realness, and spatial presence experienced by participants. Three items (SP2, INV3, REAL1) were dropped due to their being ambiguous and not adding substantially to the scale’s reliability. An example of the used items is “I had a sense of acting in the virtual space, rather than operating something from outside.” The total score was determined by calculating the mean score across all items. The questionnaire items were scaled, following a recent evaluation study (Melo et al., 2023), with scores ranging from 0 (*low presence*) to 6 (*high presence*). The internal consistency of the scale was determined as  $\alpha = 0.74$ .

Different types of *cognitive load* were assessed with a scale developed by Eysink et al. (2009). Extraneous load was surveyed with three items, including “How easy or difficult is it for you to work with the learning environment?” Intrinsic load was surveyed with the item “How easy or difficult do you consider the topic bridges and their characteristics?” Germane load was measured with the item, “How easy or difficult was it to understand the virtual reality learning environment?” Scale anchors ranged from 1 (*low cognitive load*) to 9 (*high cognitive load*). The three items of the extraneous load scale were aggregated to a mean value.

*System usability* was measured with a 10-item scale by Brooke (1996). The scale measures general usability and has been used in many usability studies. An example item is “I think that I would like to use this virtual reality learning environment frequently.” Item scores were transformed and summed following the test manual to a total score that ranged between 0 (*low system usability*) and 100 (*perfect usability*). The scale reached an internal consistency of  $\alpha = 0.69$ .

*Satisfaction* was assessed with an eight-item scale developed by Chou and Liu (2005). The scale focuses on learning satisfaction in technology-enhanced educational contexts. Here is a sample item included in the scale: “I was satisfied with the virtual reality learning environment.” Likert-type response anchors from 1 (*not agree at all*) to 5 (*fully agree*) were used. One item was formulated negatively and reverse-scored during data preparation. We obtained the total satisfaction score by averaging the scores for all scale items. The reliability of this scale was  $\alpha = 0.81$ .

### 2.4. Semi-structured interviews

The interviews focused on the topics of usability and satisfaction. They were conducted in German following a semi-structured interview protocol consisting of three questions: 1) “What did you think about the interaction with the virtual reality environment?” 2) “How user-friendly was the virtual reality environment?” and 3) “How valuable and beneficial was the virtual reality environment?” When replies were vague or participants did not seem to understand the question, we rephrased the question. Participants’ utterances were recorded on tape and transcribed. All responses were then categorized as pertaining to either system usability or satisfaction. Here, system usability is defined as the evaluation of the usability of the learning environment, with descriptions of purely technical aspects explicitly not taken into account. Satisfaction refers to the congruence among user expectations, usage requirements, and user experience. Depending on our interpretation of the content, for each categorized utterance, a *low* (1), *medium* (2), or *high* (3) score was assigned. For instance, the response “User friendly. it was very quick and easy” (Participant 1) was categorized under system usability and assigned a high score. Two raters coded all eight interviews and resolved disagreements on codes and scores until full agreement was reached. For analyzing the interview data, we calculated relative and absolute frequencies and mode values. MAXQDA (VERBI Software, 2023) was used for transcription and categorization. Microsoft Excel and R version 4.1.3 (R Core Team, 2023) were used for data analysis.

### 2.5. Procedure, experimental design, and execution

The study consisted of a pre-survey, a digital tutorial, a learning phase, and a post-survey. The pre-survey took only about 2 min to complete and was mainly focused on demographics and prior experience with VR. The digital tutorial had a duration of 4 min. It familiarized the participants with all types of user interaction included in the study and prompted them to practice these interactions. Moreover, it explained to the participants which tasks they had to accomplish in the learning phase. During the learning phase, participants learned about simplified construction engineering topics through a VR field trip. This phase lasted 7 min, so participants had sufficient time to experience the bridge and gather information through the interactive elements. In the post-survey, participants completed surveys on presence, cognitive load, ease of use, usefulness, system usability, and satisfaction. This phase did not have a time limit, but most participants finished it in 15 min. Afterward, a semi-structured interview was conducted with a subsample of the participants. Depending on the length of the answers, this interview had a duration ranging between 5 and 9 min. There were no breaks during the study, and all experimental parts occurred sequentially.

As this study also explores the topic of perceived authenticity, a two-group between-subjects design was used in the learning phase (see Fink et al., 2023a for more details). One group learned from a VR learning environment, including a bridge model created through 3D modeling with a graphics program (3D-model group). The other group received a VR learning environment comprising a bridge model created by photogrammetry (photogrammetry group)—a technology that generates from sensor data lifelike digital models. As system usability, satisfaction and variables connected to usability did not differ significantly between the two groups (see Supplementary Table 1), data from both groups were combined for this paper’s analyses. A screenshot of both learning environments—which were completely identical apart from the bridge object—is displayed in Fig. 1 [C].

The study was conducted in a VR laboratory under the supervision of a psychologically and methodologically trained experimenter. The experimenter closely followed an experimenter’s guide through all steps of the study. Participants wore an HTC Vive as an HMD connected to a gaming laptop via cable. The described VR learning environment was run on the gaming laptop as a Windows application. Participants chose teleportation platforms and selected buttons with a controller using a

virtual laser pointer.

## 2.6. Statistical analyses and power analyses of quantitative analyses

Quantitative data were analyzed with R version 4.1.3 (R Core Team, 2023). RQ1 and RQ2 were investigated with one-tailed, single-sample *t*-tests against a hypothesized value. For RQ3, RQ4, and the explorative analysis predicting ease of use with cognitive load, multiple linear regressions were conducted. Before carrying out the regressions, the assumption of independent errors was checked with Durbin-Watson tests. The data for RQ3 (Durbin-Watson value = 1.98,  $p = .484$ ), RQ4 (Durbin-Watson value = 2.08,  $p = .641$ ), and the explorative analysis (Durbin-Watson value = 2.20,  $p = .539$ ), met the assumptions of independent errors. Due to the frequently observed medium correlation among cognitive load types (Kriegelstein et al., 2023), all three types of cognitive load were entered together in regression models. The alpha level was set to  $p < .05$  for all statistical analyses.

A priori power analyses were conducted using the program G\*Power (Faul et al., 2014). For RQ1 and RQ2, power analyses were conducted for one-tailed, single sample *t*-tests against a constant value, medium effects ( $d = 0.50$ ), an alpha error level of 5%, and a power of 80%. The power analyses determined a necessary sample size of 27 participants for these research questions. For the power analyses of the regression models of RQ3 and RQ4, medium to large effects ( $f^2 = 0.25$ ) were assumed. We calculated statistics with six predictors, an alpha error level of 5%, and a power of 80%. The power analyses yielded a sample estimate of at least 62 participants.

## 3. Results

### 3.1. Frequencies, descriptives, and correlational analyses

The absolute and relative frequencies of the qualitative analyses of the semi-structured interviews are reported in Table 1. The frequencies suggest that utterances tagged with high usability and satisfaction scores were more frequent than those with medium and low scores.

Descriptive statistics for the surveys are reported in Table 2, along with intercorrelations. Usefulness, ease of use, system usability, and satisfaction attained, on a descriptive level, relatively high means. Presence scores were medium. All three types of cognitive load were rated low to medium. With respect to intercorrelations, system usability was associated with usefulness, ease of use, presence, and all cognitive load types. Satisfaction was found to be correlated with usefulness, presence, extraneous load, and intrinsic load. Moreover, the different cognitive load types were found to be related to each other. Ease of use correlated negatively with extraneous load and germane load.

### 3.2. System usability and satisfaction in the VR learning environment (RQ1 and RQ2)

RQ1 pertains to system usability in the VR learning environment. The qualitative data from the interviews presented in Table 1 show that, in sum, 92% of utterances about usability were tagged with a medium or high score, and only 8% were categorized with a low score. Moreover, high scores formed the majority (57%). These numbers highlight that participants mainly had positive experiences with usability. Further, some key interview statements also add support to these numbers. One

participant reported, “I actually got along really well with the user interface and the HMD. I thought it would be more complicated” (Participant 5). Other participants found that navigation with teleportation was “intuitive and good. Particularly because it reduced motion sickness” (Participant 8) and that interaction with the user interface “could be quickly grasped through the tutorial” (Participant 3). Moreover, several participants found VR useful for illustrating learning content that is challenging to imagine (e.g., blueprints of architectural objects). Quantitative analyses of the system usability data verified these results. A one-sample *t*-test supported H1, that a more than acceptable level of system usability (scale mean  $> 70$ ) was achieved ( $t(63) = 8.05$ ,  $d = 1.01$ ,  $p < .001$ , see Fig. 2 [A]).

RQ2 concerns how satisfied participants were with the VR learning environment. Table 1 indicates that 88% of utterances about satisfaction were tagged with a medium or high score; just 12% were categorized with a low score. In addition, high scores again formed the majority of responses (69%). These results suggest that participants were satisfied with the learning environment. Select interview statements also substantiate the above conclusion. A participant said they were satisfied with the learning environment because it “was more effective than reading about the same topic due to using two channels (auditive and visual) as well as interaction” (Participant 1). Other participants reported that “the provided 3D images ... fostered an understanding of the topic” (Participant 2); they could “learn actively” (Participant 8); and “focus to 100% on what they were doing” (Participant 3). Two aspects participants were not satisfied with should also be mentioned. One participant reported feeling eye strain (Participant 8), and another thought it was difficult to distinguish what was relevant in the environment (Participant 7). Nevertheless, the overall positive satisfaction evaluation aligns with the quantitative data. A one-sample *t*-test substantiated H2, that a more than acceptable level of satisfaction (scale mean  $> 3.0$ ) was achieved ( $t(63) = 15.87$ ,  $d = 1.98$ ,  $p < .001$ , see Fig. 2 [B]).

### 3.3. Connected variables of system usability (RQ3)

To address RQ3, we conducted a linear regression with system usability as a criterion and connected variables as predictors (see Table 3). The regression equation was significant, and the model explained high amounts of variance ( $F(6, 57) = 20.23$ ,  $p < .001$ ,  $R^2 = 0.68$ ). Ease of use, usefulness, presence, and extraneous load all predicted system usability as hypothesized (H3.1, H3.2, H3.3, H3.4).

### 3.4. Connected variables of satisfaction (RQ4)

To gain insights into RQ4, we conducted a linear regression with satisfaction as the outcome and connected variables as explanatory variables (see Table 4). The regression equation was significant, and the model explained medium amounts of variance ( $F(6, 57) = 7.46$ ,  $p < .001$ ,  $R^2 = 0.44$ ). Usefulness, presence, and intrinsic load all predicted satisfaction, supporting our hypotheses (H4.2, H4.3, and H4.4). However, ease of use did not predict satisfaction, contrary to H4.1.

### 3.5. Explorative analysis of cognitive load and ease of use

Inspecting the intercorrelations in Table 2, we discovered medium negative correlations between ease of use and extraneous and germane cognitive load. To further examine these intercorrelations, we conducted a multiple linear regression to predict ease of use based on all three types of cognitive load (see Table 5). The regression equation was significant and predicted small amounts of variance ( $F(3, 60) = 5.29$ ,  $p = .003$ ,  $R^2 = 0.21$ ). Extraneous cognitive load was determined to be the only significant predictor of ease of use.

**Table 1**  
Frequency statistics on the semi-structured interviews.

Variable	Low Score	Medium Score	High Score
Usability	3 (8%)	13 (35%)	21 (57%)
Satisfaction	2 (12%)	3 (19%)	11 (69%)

*Note.* Absolute and relative frequencies of the scores. The scores had the possible values of low (1), medium (2), or high (3).

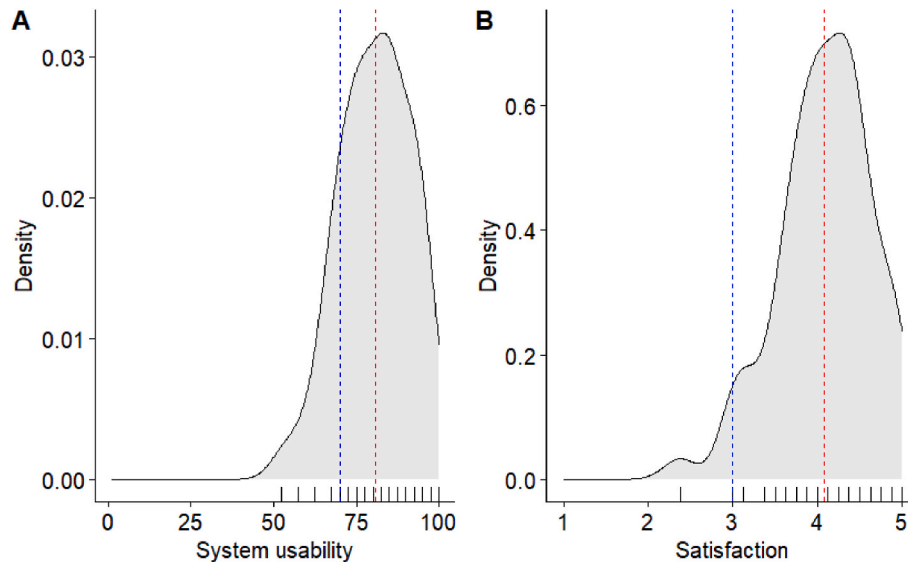
**Table 2**

Descriptive statistics and intercorrelations of the survey variables.

Variable	<i>M</i> ( <i>SD</i> )	1.	2.	3.	4.	5.	6.	7.
1. Usefulness <sup>a</sup>	3.94 (0.72)	–						
2. Ease of use <sup>a</sup>	4.57 (0.48)	.05	–					
3. Presence <sup>b</sup>	3.76 (0.83)	.10	.06	–				
4. Extraneous load <sup>c</sup>	2.84 (1.18)	–.18	–.42***	–.17	–			
5. Intrinsic load <sup>c</sup>	3.78 (1.72)	–.12	–.15	–.17	.36**	–		
6. Germane load <sup>c</sup>	1.80 (1.21)	–.07	–.36**	–.24	.46***	.30*	–	
7. System usability <sup>d</sup>	80.70 (10.64)	.34**	.61***	.44***	–.57***	–.35**	–.44***	–
8. Satisfaction <sup>a</sup>	4.08 (0.55)	.38**	.19	.49***	–.32*	–.36**	–.18	.47**

Note. Two-tailed Pearson correlations of the variables. <sup>a</sup> Range from 1 (low) to 5 (high). <sup>b</sup> Range from 0 (low) to 6 (high). <sup>c</sup> Range from 1 (low) to 9 (high).

<sup>d</sup> Range from 0 (low) to 100 (high). \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

**Fig. 2.** Density plots on system usability and satisfaction

Plot [A] contains system usability scores and plot [B] satisfaction scores. The blue line shows the contrasted threshold, the red line is the observed sample mean.

**Table 3**

Regression analysis with system usability as a criterion.

Predictor	<i>b</i>	<i>b</i> 95% CI [LL, UL]	$\beta$	$\beta$ 95% CI [LL, UL]	<i>p</i>
Intercept	16.08	[–7.62, 39.77]			.180
Ease of use***	9.83	[6.12, 13.55]	0.45	[0.28, 0.62]	<.001
Usefulness**	3.38	[1.10, 5.65]	0.23	[0.07, 0.38]	.004
Presence***	4.12	[2.13, 6.11]	0.32	[0.17, 0.48]	<.001
Extraneous load*	–2.07	[–3.75, –0.39]	–0.23	[–0.42, –0.04]	.017
Intrinsic load	–0.66	[–1.67, 0.35]	–0.11	[–0.27, 0.06]	.194
Germane load	–0.42	[–1.99, 1.14]	–0.05	[–0.23, 0.13]	.589

Note. *b* represents unstandardized regression weights.  $\beta$  represents standardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

#### 4. Discussion

##### 4.1. System usability and satisfaction in the VR learning environment (RQ1–RQ2)

The first research question addresses the level of system usability in the VR learning environment. Participants reported in their interviews that the usability levels were so high that they found interacting with the system intuitive and seamless. Two main reasons for the high levels of system usability achieved can be inferred from the interview statements. First, the learning environment only included simple user interactions (teleportation, video screens, and visualizations) and few overly

**Table 4**

Regression analysis with satisfaction as a criterion.

Predictor	<i>b</i>	<i>b</i> 95% CI [LL, UL]	$\beta$	$\beta$ 95% CI [LL, UL]	<i>p</i>
Intercept*	1.97	[0.36, 3.58]			.017
Ease of use	0.12	[–0.14, 0.37]	0.10	[–0.12, 0.33]	.358
Usefulness**	0.23	[0.07, 0.38]	0.30	[0.10, 0.50]	.005
Presence***	0.27	[0.14, 0.41]	0.42	[0.21, 0.62]	<.001
Extraneous load	–0.05	[–0.17, 0.06]	–0.11	[–0.36, 0.14]	.376
Intrinsic load*	–0.07	[–0.14, –0.00]	–0.23	[–0.45, –0.02]	.036
Germane load	0.04	[–0.06, 0.15]	0.10	[–0.14, 0.33]	.421

Note. *b* represents unstandardized regression weights.  $\beta$  represents standardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

complex features. Second, a digital tutorial was provided in which participants practiced all included interactions. The quantitative analyses verified the positive evaluation from the interviews: System usability levels were relatively high and can be considered more than acceptable (H1). Our usability values were higher compared to corresponding values in two studies on complex VR learning environments (Huang et al., 2021; Othman et al., 2022). Thus, our results indicate that complex VR learning environments created by researchers can achieve appropriate usability levels equal to those of mobile applications and learning with multimedia (Vlachogianni & Tselios, 2022). When learning environments reach such high usability scores, long-term



**Table 5**  
Regression analysis with ease of use as a criterion.

Predictor	<i>b</i>	<i>b</i> 95% CI [LL, UL]	$\beta$	$\beta$ 95% CI [LL, UL]	<i>p</i>
Intercept***	5.08	[4.75, 5.41]			<.001
Extraneous load*	−0.13	[−0.24, −0.02]	−0.32	[−0.59, −0.06]	.019
Intrinsic load	0.01	[−0.06, 0.08]	0.03	[−0.22, 0.28]	.807
Germane load	−0.09	[−0.19, 0.02]	−0.22	[−0.48, 0.04]	.099

Note. *b* represents unstandardized regression weights.  $\beta$  represents standardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit. \*\*\**p* < .001, \*\**p* < .01, \**p* < .05.

adoption and continued use of VR become more likely (Nagy, 2018).

To answer the second research question, we examined participants' satisfaction with the VR learning environment. Qualitative analyses of the interviews showed that participants were, in general, satisfied. The quantitative analyses also verified that participants experienced a more than acceptable level of satisfaction (H2), consistent with high satisfaction scores reported in other VR studies (Chang et al., 2022; Chao et al., 2023; Makransky et al., 2019; Yang et al., 2022). We can conclude from these quantitative and qualitative results that university students generally regard VR field trips as an appropriate instructional method. This point is essential, because VR is extensively used to replace excursions, which can be expensive and require much planning and time (Garcia et al., 2023). Adding to the literature on engineering education (Dos Anjos et al., 2021), our study also highlights that VR can elicit high levels of satisfaction when teaching simplified engineering topics to university students of all majors. Furthermore, the qualitative interviews allow us to draw conclusions about the characteristics and processes that positively influence user satisfaction. Participants particularly appreciated the visualizations included and the high level of interactivity in the learning environment. Moreover, they mentioned that they had a multi-sensory experience and a high level of engagement. Based on these findings, future studies should explore how various characteristics of the VR learning environments (e.g., navigation and scaffolding types, see Fink et al., 2023b) and the affective and cognitive processes experienced (e.g., feelings of enjoyment, see Makransky & Lilleholt, 2018) relate to satisfaction.

#### 4.2. Connected variables of system usability (RQ3)

To approach the third research question, we investigated variables connected with system usability. Similar to assumptions included in the TAM (e.g., Davis, 1989), ease of use and usefulness were found to explain system usability (H3.1 and H3.2), and the effect sizes of their relations were medium and small, respectively. The effect sizes are comparable in magnitude to relations found for these predictors and other usability facets, such as attitude toward using and behavioral intention, in the regular e-learning literature (Sumak et al., 2011). Our results add to the literature that ease of use and usefulness predict usability facets also in complex VR learning environments experienced with HMDs. This finding is significant, because data on the association of ease of use and usefulness with system usability were missing for this specific context (Vlachogianni & Tselios, 2022). It should be noted that a direct comparison of our results on system usability with findings from other theories of technology acceptance, such as UTAUT (Venkatesh et al., 2003) and its specific adoption for VR contexts (Ustun et al., 2023), would be complicated. These technology acceptance theories are more extensive and incorporate ease of use and usefulness as components of larger scales.

Next, let us discuss the association between presence and system usability. As expected, presence explained variance in system usability (H3.3). This result aligns with studies using different types of VR, including HMDs and CAVE systems (Brade et al., 2017; Voinescu et al.,

2023; Wienrich et al., 2018). One explanation of this relationship could be that higher presence feelings increase system usability through positive emotions (Brade et al., 2017). Another explanation could be that a higher feeling of presence may be associated with improved attention or concentration, which could, in turn, positively affect participants' usability experience (Voinescu et al., 2023; Wienrich et al., 2018; Witmer & Singer, 1998). This mechanism has been well established in the literature, but so far, only one VR study appears to have examined it closely: In this study by Voinescu et al. (2023), participants completed an attention task in VR with HMDs and rated system usability and presence. Presence predicted system usability, but was—against the described mechanism—not related to attention performance (Voinescu et al., 2023). In brief, our study provides additional evidence for the relationship between presence and usability. Explanations for this relationship need to be further explored by future researchers.

Lastly, we discuss the relation between cognitive load and system usability. Before interpreting our results, let us briefly inspect the cognitive load level. Since all three cognitive load scores were low to medium, we can conclude that cognitive overload was not an issue in our study. In line with hypothesis H3.4, extraneous load proved a significant negative predictor in our multiple regression. This finding suggests that extraneous cognitive load can reduce the usability of VR learning environments, as proposed in the literature (Hollender et al., 2010). However, it should be noted the effect size of this predictor was relatively small ( $\beta = -0.23$ ), and further replication studies are necessary to support this result. Our results do not indicate relationships of intrinsic and germane load with usability. The effect sizes of intrinsic and germane load in the multiple regression were visibly smaller than that for extraneous load and not close to the 5% alpha level. These results are consistent with a conceptualization of intrinsic and germane load independent from the system usability of the learning environment (Hollender et al., 2010). This conceptualization makes sense because intrinsic and germane load arise mainly from the demands of the learning material and schema construction processes (Sweller et al., 1998).

#### 4.3. Connected variables of satisfaction (RQ4)

The fourth research question focuses on variables connected with satisfaction. Contrary to our expectations, ease of use did not predict satisfaction (H4.1). We assumed this variable would be associated with satisfaction based on the following reasoning. Satisfaction is part of the ISO usability definition (ISO, 2018), and various usability variables (e.g., attitudes toward using) are related to ease of use (Sumak et al., 2011). Our results indicate that the known relations between ease of use and usability do not extend to satisfaction. A conceptualization of satisfaction independent from usability as a separate perceived learning outcome (Makransky & Lilleholt, 2018) would fit these results. Such perceived learning outcomes can be predicted, in particular, by positive affective and motivational states, but are not theorized to be related to usability variables (Makransky & Lilleholt, 2018). That being said, the conceptualization of satisfaction we selected based on ISO (2018) should also not be viewed too critically, because other prominent usability conceptualizations also include satisfaction as a key component (Harrison et al., 2013; Nielsen, 1994).

Moreover, we found that usefulness positively predicted satisfaction (H4.2). This result corresponds to findings from a meta-analysis on e-learning, which indicates a link between usefulness and satisfaction in four out of five studies (Sumak et al., 2011). Likewise, Ohk et al. (2015) discovered a link between usefulness and satisfaction when using a smartphone application. Our study and the studies above demonstrate that the value ascribed to a system affects the satisfaction experienced with it. Further, usefulness has some similarities to value judgments in motivational theories. According to Eccles (1983), positive value judgments arise when learning environments are pleasant (*intrinsic value*), relevant to the user's identity (*attainment value*), and relate to the future



(utility value). These factors could be considered when developing VR learning environments to ensure high usefulness. We also examined the association of satisfaction with presence and found presence to be a positive predictor of satisfaction (H4.3), suggesting that higher presence contributes to the experience of greater learning satisfaction. This result aligns with the theoretical argument that increased flow experiences and similar states like presence positively impact satisfaction (Liu et al., 2023). As far as we know, only one study has investigated this relationship in the context of a desktop-based VR application (Liu et al., 2023). Since this study used other questionnaires to measure presence and usability, and our VR was conducted with HMDs, our study represents a vital contribution to the literature. Due to the described relationship between presence and satisfaction, educational designers and developers should aim to enhance the presence experienced by participants. Higher levels of presence might be attained when learning environments fit well with participants' values, offer a high level of interactivity, are visually realistic, and represent the critical functions of the situation they simulate (Chernikova et al., 2020; Skulmowski et al., 2021; Witmer & Singer, 1998).

Finally, we investigated the association between cognitive load and satisfaction. Based on the low to medium scores obtained for cognitive load, we can assume that the learning environment did not lead to cognitive overload. Intrinsic cognitive load was found to be a negative predictor of satisfaction (H4.4), implying that the high cognitive load created by the content decreased learning satisfaction. We expected this finding, assuming that participants' consideration of the material as challenging may frustrate them and thus decrease their satisfaction (Kalyuga, 2011). Nevertheless, this result seems noteworthy, because it occurred in a complex VR learning environment with only low to medium cognitive load. Extraneous and germane load, as assumed, did not predict satisfaction.

#### 4.4. Explorative analysis of cognitive load and ease of use

Through our explorative multiple regression, we found that extraneous cognitive load negatively predicted ease of use, while intrinsic and germane cognitive load did not. This finding adds to research discussing the dependence of cognitive load and usability variables (e.g., Hollender et al., 2010). The observed prediction of ease of use by extraneous load demonstrates that not only system usability, but also *variables connected to usability* can be related to extraneous load. Consequently, the interrelations between cognitive load types and variables connected to usability should be researched further.

#### 4.5. Limitations

One limitation of our study concerns the measures used. Instead of measuring usability variables extensively with larger surveys (e.g., UTAUT from Venkatesh et al., 2003, or TAM 3 from Venkatesh & Bala, 2008), we used several short measures for system usability, satisfaction, ease of use, usefulness, presence, and cognitive load. With this approach, several components of the mentioned larger surveys were skipped and thus could not be examined. For instance, TAM3 contains a set of individual variables (e.g., computer self-efficacy, computer playfulness, computer anxiety) and externally influenced variables (e.g., the voluntariness of system use, the system's image, and job relevancy) associated with acceptance. Likewise, usability is broadly operationalized in models such as the TAM3 with many items and subcomponents that do not entirely fit the ISO (2018) usability definition. Our approach of using shorter scales and sticking with the ISO usability definition had the advantage that we were able to analyze variables such as cognitive load and presence that are not contained in larger surveys, but are theoretically relevant for learning in VR.

Another limitation of our study concerns the relatively simple type of user interaction in the VR design. Participants used a laser pointer to teleport across platforms and to select buttons. However, more advanced

types of interaction can be implemented in VR. In some learning environments, participants can navigate freely by walking (Schmeil et al., 2012). Other learning environments enable participants to interact directly with the user interface by pushing buttons or by manipulating and using objects with hand motions (Schmeil et al., 2012). Because our study was focused on a learning environment with simple interaction types, our results cannot be generalized to learning environments with more complex interaction types without restrictions.

A final limitation concerns the effects of biases on our study. Possible sources of bias may be that participants felt observed and changed their behavior due to participating in a study ("Hawthorne effect") or adjusted their behavior to our expectations ("Rosenthal effect"). Let us consider here again, briefly, that the study was conducted strictly according to an experimenter's guide by psychologically and methodologically trained experimenters. Because standardized surveys were used and participants filled out surveys unobserved and anonymously, it is unlikely that the results from the survey data were affected substantially by the described biases. However, the semi-structured interviews conducted after the surveys were more susceptible to such biases (Bergen & Labonté, 2020). In these semi-structured interviews, we established rapport and prompted participants to answer to the best of their knowledge. Then, we asked questions in a pre-determined sequence. Like in any interview study, it is possible that participants answered the interview questions in a socially desirable manner or inferred our expectations of them to some extent.

#### 4.6. Practical implications and recommendations for educational design

Currently, educational VR with HMDs is rarely used in higher education. Various factors may be responsible for this. For instance, only a few educational VR applications with clear learning goals relevant to the curriculum are available in content stores, and HMDs are still expensive and require maintenance. Moreover, there is a lack of supporting materials that would enable instructors to easily and effectively integrate VR applications into their teaching. Research on other educational technologies, including instructional videos, tablets, and smartphones, has shown that usability ratings predict actual usage and long-term adoption (Alsharida et al., 2021; Nagy, 2018). Against this backdrop, our study's positive evaluations of system usability and satisfaction suggest a high potential for long-term adoption of educational VR. However, research on smartphone adoption indicates that long-term technology adoption also depends on many other factors, including the popularity and price of devices and applications and the function of signaling technological affinity within a social group (Kim & Kim, 2014). These points illustrate that establishing VR in higher education requires collaboration among multiple parties with clear learning and assessment goals, a sensible budget, and a realistic timeline.

Our results also have implications for assessing the technology acceptance of VR. The UTAUT for VR contexts (Ustun et al., 2023) distinguishes between the subscales performance expectancy, social influence, effort expectancy, and facilitating conditions. This instrument seems appropriate for investigating the technology acceptance of VR among university students in the coming years. Our study has shown that cognitive load and presence are linked to system usability and satisfaction. These associations suggest that it may be beneficial to supplement technology acceptance scales for VR by measuring critical variables included in frameworks for VR learning. Behavioral intention and usage may be better predicted if technology acceptance models and VR-specific variables are used in combination.

Our study also has implications for designing virtual field trips. System usability scores were high and surpassed scores in two other educational VR studies (Huang et al., 2021; Othman et al., 2022). This effect can mainly be attributed to using simple UI and providing participants with a detailed tutorial. Consequently, we recommend these two steps to others implementing VR field trips in educational contexts. Satisfaction was also high, indicating that participants enjoyed the

virtual field trip. We believe that the high satisfaction was achieved by including powerful visualizations and a high level of interactivity. Therefore, researchers developing virtual field trips should implement visualizations and interactivity as best as possible. This point leads us to an issue for which virtual field trips can be criticized. Virtual field trips are often conducted like real field trips and have many phases where learners passively observe. Consistent with the ICAP framework (Chi & Wylie, 2014), active, constructive, and interactive learning are beneficial. We recommend that scaffolding (e.g., quizzes and self-explanation prompts), game elements (e.g., points and leaderboards), and collaboration opportunities (e.g., with a virtual agent or a peer) are incorporated into virtual field trips to increase their effectiveness and avoid passive learning.

In addition, further recommendations for educational design can be provided. Our study shows that the connected variables of system usability and satisfaction differ. System usability is affected by ease of use, usefulness, presence, and extraneous load. Satisfaction is linked to usefulness, presence, and intrinsic load. Educational designers and developers should, thus, keep the following recommendations in mind: System usability can be optimized through increasing ease of use, usefulness, and presence and reducing extraneous load. Satisfaction can be enhanced by maximizing usefulness and presence and preventing an overly high level of intrinsic load. Moreover, our study reveals that extraneous cognitive load negatively predicts ease of use. This result suggests that decreasing extraneous load may also have positive effects on connected variables to usability. The interdependence of cognitive load and usability variables (Hollender et al., 2010) should be explored further to provide more guidance to practitioners.

#### 4.7. Summary and conclusion

We investigated usability and satisfaction in a VR learning environment. Overall, the system usability was evaluated positively and achieved levels equal to mobile applications and learning with multimedia. This result signals that VR learning environments have a high potential for long-term adoption, although this process may take time and depend on a range of other factors. Satisfaction was considered sufficient, indicating that learners perceive VR field trips as a valuable didactical format. As VR field trips are increasingly used, we discussed how to effectively design this format. We recommend incorporating scaffolding, game elements, and collaboration opportunities into virtual field trips to increase their effectiveness and avoid passive learning. In addition to these topics, this study examined relations among system usability, satisfaction, and connected variables. Ease of use, usefulness, presence, and extraneous load all predicted system usability. Usefulness, presence, and intrinsic load explained variance in satisfaction. These results indicate that the relationships between system usability and satisfaction with connected variables differ. The following recommendations can serve as a guide for educational designers: System usability can be optimized by maximizing ease of use, usefulness, and presence and minimizing extraneous load. Satisfaction can be improved by increasing usefulness and presence while holding the intrinsic load in check.

#### 5. Statements on open data and ethics

The data can be obtained upon request by contacting the corresponding author. The study involving human participants was reviewed and approved by ethics committee of Universität der Bundeswehr München. The participants provided their written informed consent to participate in this study.

#### Author contributions

**Maximilian C. Fink:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project

administration. **Volker Eisenlauer:** Conceptualization, Writing – Review & Editing, Resources. **Bernhard Ertl:** Conceptualization, Methodology, Formal analysis, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

#### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL in order to translate and improve writing of passages. After using this service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This research paper is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr [project RISK.twin]. dtec.bw is funded by the European Union – NextGenerationEU. We thank Diana Sosa and Lukas Hart for their support in creating the photogrammetry-based learning environment and their assistance in data collection. MF thanks Larissa Kaltefleiter for inspiring discussions on the topic.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cexr.2023.100043>.

#### References

- Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, 166, Article 104154.
- Almalki, S. (2016). Integrating quantitative and qualitative data in mixed methods research – challenges and benefits. *Journal of Education and Learning*, 5(3), 288. <https://doi.org/10.5539/jel.v5n3p288>
- Alsharida, R. A., Hammood, M. M., & Al-Emran, M. (2021). Mobile learning adoption: A systematic review of the technology acceptance model from 2017 to 2020. *International Journal of Emerging Technologies in Learning*, 16, 147. <https://doi.org/10.3991/ijet.v16i05.18093>, 05.
- Ardito, C., Costabile, M. F., Marsico, M. de, Lanzilotti, R., Levialdi, S., Roselli, T., & Rossano, V. (2006). An approach to usability evaluation of e-learning applications. *Universal Access in the Information Society*, 4(3), 270–283. <https://doi.org/10.1007/s10209-005-0008-6>
- Bergen, N., & Labonté, R. (2020). “Everything Is perfect, and we have no problems”: Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research*, 30(5), 783–792. <https://doi.org/10.1177/1049732319889354>
- Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., & Klimant, P. (2017). Being there again – presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. *International Journal of Human-Computer Studies*, 101, 76–87. <https://doi.org/10.1016/j.ijhcs.2017.01.004>
- Brooke, J. (1996). Sus - a quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. Lyall, & B. W. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Chang, C.-Y., Sung, H.-Y., Guo, J.-L., Chang, B.-Y., & Kuo, F.-R. (2022). Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interactive Learning Environments*, 30(3), 400–416. <https://doi.org/10.1080/10494820.2019.1661854>
- Chao, Y.-P., Kang, C.-J., Chuang, H.-H., Hsieh, M.-J., Chang, Y.-C., Kuo, T. B. J., Yang, C. C. H., Huang, C.-G., Fang, T.-J., Li, H.-Y., & Lee, L.-A. (2023). Comparison of the effect of 360° versus two-dimensional virtual reality video on history taking and physical examination skills learning among undergraduate medical students: A randomized controlled trial. *Virtual Reality*, 27(2), 637–650. <https://doi.org/10.1007/s10055-022-00664-0>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>

- Chou, S.-W., & Liu, C.-H. (2005). Learning effectiveness in a web-based virtual learning environment: A learner control perspective. *Journal of Computer Assisted Learning*, 21(1), 65–76. <https://doi.org/10.1111/j.1365-2729.2005.00114.x>
- Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Dos Anjos, F. E. V., Rocha, L. A. O., Da Silva, D. O., Pacheco, R., & Pinheiro, D. M. B. (2021). Impacts of the application of virtual and augmented reality on teaching-learning processes in engineering courses. *International Journal of Virtual and Personal Learning Environments*, 12(1), 1–19. <https://doi.org/10.4018/IJVPLE.291541>
- Eccles, J. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *A series of books in psychology. Achievement and achievement motives: Psychological and sociological approaches*. W.H. Freeman.
- Eysink, T. H. S., de Jong, T., Berthold, K., Kolloff, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal*, 46(4), 1107–1149. <https://doi.org/10.3102/0002831209340235>
- Faul, F., Buchner, A., Erdfelder, E., & Lang, A. G. (2014). *G\*Power [computer software]*. <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html>
- Fink, M. C., Sosa, D., Eisenlauer, V., & Ertl, B. (2023a). Authenticity and interest in virtual reality: Findings from an experiment including educational virtual environments created with 3D modeling and photogrammetry. *Frontiers in Education*, 8, Article 969966. <https://doi.org/10.3389/educ.2023.969966>
- Fink, M. C., Eisenlauer, V., Frischbier, D., & Ertl, B. (2023b). Zentrale Merkmale immersiver VR-Lernumgebungen: Eine Taxonomie veranschaulicht anhand von drei Beispielen. In B. Zinn (Ed.), *Virtual Reality, Augmented Reality und Serious Games als Educational Technologies in der Beruflichen Bildung* (pp. 13–43). Steiner Verlag.
- Garcia, M. B., Nadelson, L. S., & Yeh, A. (2023). “We’re going on a virtual trip!”: A switching-replications experiment of 360-degree videos as a physical field trip alternative in primary education. *International Journal of Child Care and Education Policy*, 17(1), 4. <https://doi.org/10.1186/s40723-023-00110-x>
- Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), 1. <https://doi.org/10.1186/2194-0827-1-1>
- Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, 26(6), 1278–1288. <https://doi.org/10.1016/j.chb.2010.05.031>
- Huang, H., Lin, C., & Cai, D. (2021). Enhancing the learning effect of virtual reality 3D modeling: A new model of learner’s design collaboration and a comparison of its field system usability. *Universal Access in the Information Society*, 20(3), 429–440. <https://doi.org/10.1007/s10209-020-00750-7>
- ISO. (2018). *Usability: Definitions and concepts*. London: ISO. <https://www.iso.org/obp/ui/#iso:std:iso:9241-11:ed-2:v1:en:sec:A>
- Jensen, L., & Konrad, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515–1529. <https://doi.org/10.1007/s10639-017-9676-0>
- Kalyuga, S. (2011). Cognitive load theory: Implications for affective computing. In R. C. Murray, & P. M. McCarthy (Eds.), *Proceedings of the twenty-fourth international Florida artificial intelligence research society conference* (pp. 105–110). AAAI Press.
- Karaoglan-Yilmaz, F. G., Yilmaz, R., Zhang, K., & Ustun, A. B. (2023). Development of educational virtual reality attitude scale: A validity and reliability study. *Virtual Reality*, 27(3), 1875–1885. <https://doi.org/10.1007/s10055-023-00778-z>
- Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A systematic review of virtual reality in education. *Themes in Science & Technology Education*, 10(2), 85–119.
- Kim, A., & Kim, K. (2014). *User experience and the multi-stage adoption of mobile apps*. Journal of Information Technology Applications & Management.
- Kriegelstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review*, 35(1). <https://doi.org/10.1007/s10648-023-09738-0>
- Lanzo, J. A., Valentine, A., Sohel, F., Yapp, A. Y. T., Muparadzi, K. C., & Abdelmalek, M. (2020). A review of the uses of virtual reality in engineering education. *Computer Applications in Engineering Education*, 28(3), 748–763. <https://doi.org/10.1002/cae.22243>
- Lee, E.-L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, 55(4), 1424–1442. <https://doi.org/10.1016/j.compedu.2010.06.006>
- Liu, Z., Yu, P., Liu, J., Pi, Z., & Cui, W. (2023). How do students’ self-regulation skills affect learning satisfaction and continuous intention within desktop-based virtual reality? A structural equation modelling approach. *British Journal of Educational Technology*, 54(3), 667–685. <https://doi.org/10.1111/bjet.13278>
- Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research & Development*, 66(5), 1141–1164. <https://doi.org/10.1007/s11423-018-9581-2>
- Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modelling approach. *Computers & Education*, 134, 15–30. <https://doi.org/10.1016/j.compedu.2019.02.002>
- 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*, 33(3), 937–958. <https://doi.org/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(11), 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Marangunic, N., & Granic, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, 14(1), 81–95. <https://doi.org/10.1007/s10209-014-0348-1>
- Melo, M., Gonçalves, G., Vasconcelos-Raposo, J., & Bessa, M. (2023). How much presence is enough? Qualitative scales for interpreting the Igroup presence questionnaire score. *IEEE Access*, 11, 24675–24685. <https://doi.org/10.1109/ACCESS.2023.3254892>
- Mikropoulos, T. A., & Bellou, J. (2006). The unique features of educational virtual reality environments. In P. Isaias, M. McPherson, & F. Bannister (Eds.), *Proceedings of the IADIS international conference e-society* (pp. 122–128). IADIS Press.
- Nagy, J. T. (2018). Evaluation of online video usage and learning satisfaction: An extension of the technology acceptance model. *International Review of Research in Open and Distance Learning*, 19(1), 159–185. <https://doi.org/10.19173/irrodl.v19i1.2886>
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Ohk, K., Park, S.-B., & Hong, J.-W. (2015). The influence of perceived usefulness, perceived ease of use, interactivity, and ease of navigation on satisfaction in mobile application. *Advanced Science and Technology Letters*, 84, 88–92. <https://doi.org/10.14257/astl.2015.84.18>
- Othman, M. K., Nogoibaeva, A., Leong, L. S., & Barawi, M. H. (2022). Usability evaluation of a virtual reality smartphone app for a living museum. *Universal Access in the Information Society*, 21, 995–1012. <https://doi.org/10.1007/s10209-021-00820-4>
- R Core Team. (2023). *R: A language and environment for statistical computing [computer software]*. Vienna, Austria <https://www.R-project.org/>
- Radiant, 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, Article 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Schmeil, A., Eppler, M., & Freitas, S. (2012). A structured approach for designing collaboration experiences for virtual worlds. *Journal of the Association for Information Systems*, 13(10), 836–860. <https://doi.org/10.17705/1jais.00309>
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10(3), 266–281. <https://doi.org/10.1162/105474601300343603>
- Skulmowski, A., Nebel, S., Remmele, M., & Rey, G. D. (2021). *Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations*. *Educational psychology review*. Advance online publication. <https://doi.org/10.1007/s10648-021-09638-1>
- Statista. (2022). *Virtual reality (VR): Digital & trends*. <https://www.statista.com/study/29689/virtual-reality-vr-statista-dossier/>
- Šumak, B., Heričko, M., & Pušnik, M. (2011). A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types. *Computers in Human Behavior*, 27(6), 2067–2077. <https://doi.org/10.1016/j.chb.2011.08.005>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Ustun, A. B., Karaoglan-Yilmaz, F. G., & Yilmaz, R. (2023). Educational UTAUT-based virtual reality acceptance scale: A validity and reliability study. *Virtual Reality*, 27(2), 1063–1076. <https://doi.org/10.1007/s10055-022-00717-4>
- Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- VERBI Software. (2023). *MAXQDA 2023 [Computer software]*. Berlin, Germany. maxqda.com.
- Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the system usability scale (SUS): A systematic review. *Journal of Research on Technology in Education*, 54(3), 392–409. <https://doi.org/10.1080/15391523.2020.1867938>
- Vogt, A. G. (2021). *Fostering deep learning in immersive virtual reality: Interplay of learner’s characteristics with internal and external support [dissertation]*. Ulm: Universität Ulm. [https://oparu.uni-ulm.de/xmlui/bitstream/handle/123456789/44276/Dissertation\\_Vogt.pdf?sequence=1&isAllowed=y](https://oparu.uni-ulm.de/xmlui/bitstream/handle/123456789/44276/Dissertation_Vogt.pdf?sequence=1&isAllowed=y)
- Voinescu, A., Petrini, K., & Stanton Fraser, D. (2023). Presence and simulator sickness predict the usability of a virtual reality attention task. *Virtual Reality*, 1–17. <https://doi.org/10.1007/s10055-023-00782-3>
- Wang, P., Wu, P., Wang, J., Chi, H.-L., & Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. *International Journal of Environmental Research and Public Health*, 15(6). <https://doi.org/10.3390/ijerph15061204>
- Wienrich, C., Döllinger, N., Kock, S., Schindler, K., & Traupe, O. (2018). Assessing user experience in virtual reality – a comparison of different measurements. In A. Marcus, & W. Wang (Eds.), *Lecture notes in computer science. Design, user experience, and usability: Theory and practice* (Vol. 10918, pp. 573–589). Springer International Publishing. [https://doi.org/10.1007/978-3-319-91797-9\\_41](https://doi.org/10.1007/978-3-319-91797-9_41)

- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240.
- Woerner, J. J. (1999). Virtual field trips in the earth science classroom. In P. A. Rubba, J. A. Rye, & P. F. Keig (Eds.), *Proceeding of the annual conference of the Association for the Education of Teachers in science. Symposium conducted at the meeting of association for the education of teachers in science*.
- Yang, H., Cai, M., Diao, Y., Liu, R., Liu, L., & Xiang, Q. (2022). How does interactive virtual reality enhance learning outcomes via emotional experiences? A structural equation modeling approach. *Frontiers in Psychology*, 13, Article 1081372. <https://doi.org/10.3389/fpsyg.2022.1081372>