




# “Feels Like an Indie Game”—Evaluation of a Virtual Field Trip Prototype on Radioactive Waste Management Research for University Education

Nico Graebling , Melanie Althaus , and Özgür Ozan Şen , *Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany*

Thomas Reimann , *Dresden University of Technology, 01062, Dresden, Germany*

Tuanny Cajuhi , *Federal Institute for Geosciences and Natural Resources, 30655, Hanover, Germany*

Gerik Scheuermann , *Leipzig University, 04109, Leipzig, Germany*

Olaf Kolditz  and Karsten Rink , *Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany*

*This article describes the design and evaluation of a virtual field trip on the topic of radioactive waste management research for university education. We created an interactive virtual tour through the Mont Terri underground research laboratory by enhancing the virtual experiment information system, designed for domain experts, with background information, illustrations, tasks, tests, and an improved user interface. To put the tour’s content into context, a conventional introductory presentation on the final disposal of radioactive waste was added. A user study with 22 participants proved a good perceived usability of the virtual tour and the virtual field trip’s ability to transfer knowledge. These results suggest a benefit of employing virtual field trips in geoscientific university courses. In addition, it is conceivable to use the virtual field trip as a tool for science communication in the context of participatory processes during nuclear waste disposal site selection processes.*

A significant part of the research on the final disposal of radioactive waste is carried out in underground research laboratories (URL). This allows scientists to study the properties of potential host rocks at the scale of an actual repository. The Mont Terri rock laboratory in Switzerland is an example of such a URL. International scientists have been studying the suitability of Opalinus Clay as a host rock for radioactive waste disposal for more than 25 years at Mont Terri.<sup>1</sup> The URL itself is not being considered

for storage, but is operated exclusively for research purposes.

Research on the final disposal of radioactive waste is relevant not only for the geoscientific community, but also for the public. For instance, in Germany, participatory processes are required by law as part of the site selection process. This ensures the possibility of involvement, which can increase social acceptance.<sup>2</sup> A precondition for participatory processes is the availability of information on the state of research to all stakeholders, as this information is needed to form an opinion. In this case, availability includes technical accessibility and a presentation in an understandable form. Scientific publications on research methods and results are not sufficient for this purpose, as they require too much prior knowledge and are not always accessible and understandable to those outside the scientific community.

---

© 2023 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

This article has supplementary downloadable material available at <https://doi.org/10.1109/MCG.2023.3328169>, provided by the authors.

Digital Object Identifier 10.1109/MCG.2023.3328169

Date of publication 27 October 2023; date of current version 25 January 2024.

We created a virtual tour through the Mont Terri URL by adapting and extending the virtual experiment information system prototype (VEIS), which we developed in previous work.<sup>3</sup> The VEIS is a 3-D tool that allows experts to view and explore heterogeneous datasets from the URL (e.g., observation data and results of numerical simulations). The virtual tour presents geological research methods and results from the URL via 3-D visualizations to make knowledge accessible to intermediates (such as students and junior researchers). Therefore, the VEIS was extended to include more contextual knowledge, explanations, illustrations, tasks, knowledge tests, and user interface (UI) features. In addition, the usability of the application was improved and educational concepts such as addressing specific learning processes according to Anderson and Krathwohl's taxonomy of cognitive learning processes were implemented for effective knowledge transfer.<sup>4</sup>

We combined the virtual tour of the Mont Terri URL with an introductory presentation to create a virtual field trip on radioactive waste management research. The focus of the presentation is set on the situation in Europe and in particular in Germany, as the evaluation of the virtual field trip took place at two German universities. The Mont Terri URL was chosen for the virtual field trip because, at the time of writing, there is no equivalent URL in Germany and German research institutions are also conducting research at the Swiss URL.

Prior work has already shown that virtual field trips can be a valuable supplement of real field trips and that the learning outcome as well as the experience might be equal or even better than on actual field trips.<sup>5,6</sup> It is nonetheless important to evaluate both, the overall concept of virtual field trips and the specific applications, as each virtual field trip and each implementation may differ in various aspects. This includes the technical setup and implementation, the application of scientific visualization concepts, and the didactic concept. All these factors play an important role concerning the usability of the application and the knowledge transfer created by the virtual field trip. Therefore, we evaluated the usability of the virtual tour and the virtual field trip's ability to transfer knowledge, even though evaluations for similar applications (e.g., Zhao et al.<sup>5</sup>, Ferro et al.<sup>7</sup>) already exist. The evaluation was conducted in two geoscientific university courses with a total of 22 participants.

In recent years, the importance of digital twins and serious games in the geoscientific domain for education and training has increased.<sup>7,8,9,10</sup> Digital twins are combinations of "a physical entity, a virtual

counterpart, and the data connections in between," which are used in a variety of applications like monitoring, simulation, and visualization.<sup>8</sup> A serious game is a game whose goal is not only to entertain but also to achieve other goals, such as gaining knowledge or supporting a decision-making process.<sup>9</sup> Depending on the actual implementation, virtual field trips for the geoscientific domain and related research fields (e.g., Dolphin et al.<sup>11</sup>, Harrington et al.<sup>12</sup>) can be considered a mixture of a visualization digital twin and a serious game.

The following section describes the adaptation of the VEIS. Then, we give details on the study design and the evaluation results. The last section presents the conclusions and an outlook to future work.

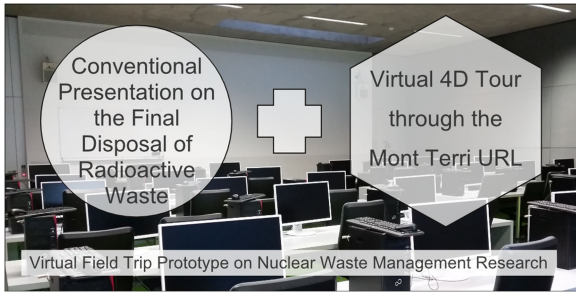
## ADAPTATION OF THE VEIS

### Development Process

We developed the virtual experiment information system prototype (VEIS, see the "The VEIS" section) in previous work with a focus on experts from the geoscientific domain. Therefore, adaptations had to be implemented before making it available to a different target group (intermediates) within this work. In addition, the application's usability had to be improved for this work, as the VEIS is a prototype. In the following, we describe the development process, including two correction cycles: First, we added new features to the VEIS in order to create the first prototype of the virtual tour. The additional features and their purposes are described in the "Additional Features" section. Second, a user experience expert carried out a heuristic evaluation (details in the "Heuristic Evaluation" section). The corrections based on this heuristic evaluation were applied in the second version of the prototype. Third, the final evaluation procedure was tested with three participants (see the "Test Evaluation" section). Their feedback was implemented in order to create the final version of the virtual tour used for the evaluations at the universities. This final virtual tour, including all additional features and improvements after the heuristic evaluation and the test evaluation, is provided as a video<sup>a</sup> in the supplemental material, available online.

Based on the taxonomy of Klippel et al.<sup>6</sup> the virtual tour alone can already be considered an *advanced virtual field trip* as it not only replicates the real laboratory but also allows for additional spatial perspectives

<sup>a</sup>Video of the final virtual tour: <https://youtu.be/kH34J9cZ3aI>

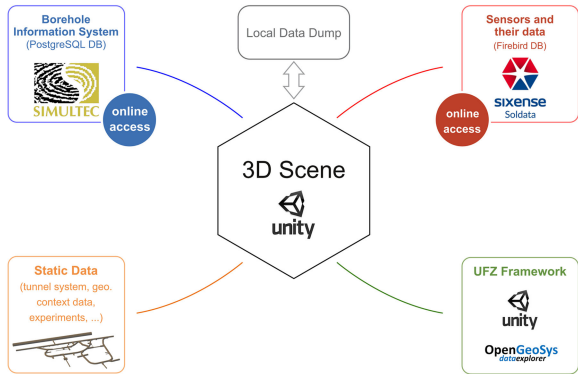


**FIGURE 1.** Two components of the virtual field trip on radioactive waste management research and in the background the computer lab at TU Dresden, where we conducted one of the evaluation sessions.

and the possibility to explore time series data and numerical simulation results. However, a conventional presentation was prepared to enhance the virtual tour with context information about the final disposal of radioactive waste because the virtual tour itself focuses specifically on the Mont Terri URL. The resulting combination of the presentation and the virtual tour, as shown in Figure 1, is what comprises our virtual field trip on radioactive waste management research.

### The VEIS

The virtual experiment information system prototype is a tool for domain experts to visually access and explore relevant research data from the Mont Terri URL with low technical barriers.<sup>3</sup> The data included in this application are depicted in Figure 2. It contains datasets of the geological context, like satellite images of the region and the stratigraphic layers surrounding the URL, as well as the tunnel system and the boreholes. Sensors measuring properties of the host rock are usually located within boreholes. The measured time series data are obtained from the URL’s productive databases and can be accessed by clicking on the representation of a borehole and selecting one of the available sensors (see Figure 3). These data are enhanced by animations of 3-D simulation results and illustrations of experiments. The VEIS contains detailed monitoring data and simulation results for three selected experiments. These vary in types of available data, domain size, time span, measured parameters, and their research objectives (see Graebing et al.<sup>3</sup> for details). The visualization of these highly heterogeneous datasets in a combined visual context supports a holistic exploration and discussions



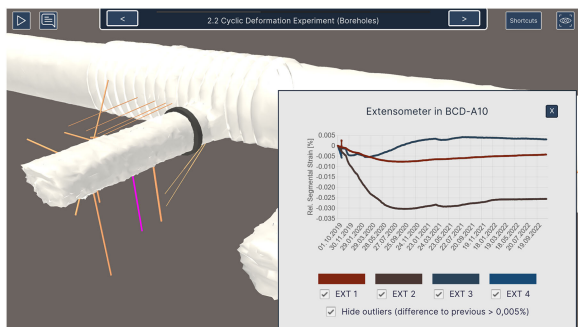
**FIGURE 2.** Data included in the original VEIS, which was used as a base for the 4-D virtual tour. (Source: Graebing et al.<sup>3</sup>; used with permission.)

among scientists. We employ the application as the basis for the virtual tour used in this work.

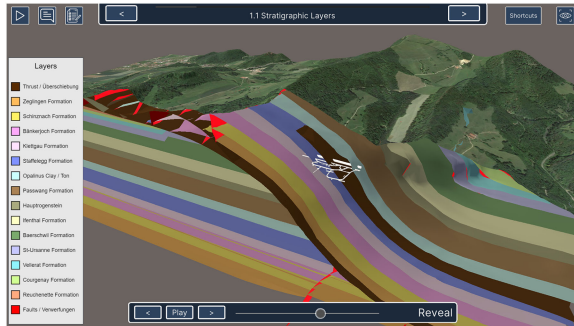
### Additional Features

We added several additional features to the VEIS in order to create the first prototype of the virtual tour.

First, the illustrations of the three experiments (see the “The VEIS” section) were extended. Their objectives and methods are less intuitive for intermediates than for experts. Therefore, we added illustrations with more context information to introduce the users to the general idea and setup of the experiments in a more descriptive and vivid way. An example for such an illustration is the introduction to the full-scale emplacement experiment. For this experiment, three heaters were placed in a tunnel to simulate the heat induced by radioactive waste and to investigate the effects of the temperature changes on the surrounding geology’s properties. Our illustration covers the



**FIGURE 3.** Accessing sensor measurements as time series data directly from the 3-D scene is a feature from VEIS that is also available in the virtual tour.



**FIGURE 4.** The tunnel system of the Mont Terri URL in its geological context. The slider at the bottom allows the user to change the stratigraphic layers' visibility.

heaters' purpose, power, and positioning within the experiment to prepare intermediates for the interpretation of the simulation results.

Second, tasks were added to six of the tour's viewpoints. A viewpoint consists of a predefined perspective (i.e., position and orientation of the virtual camera), an audio or text comment, the visibilities of datasets, and possible interactions. The tasks were designed based on the learning objectives defined in Anderson and Krathwohl's taxonomy,<sup>4</sup> which is a revision of Bloom's taxonomy of educational objectives.<sup>13</sup> Users are motivated to "remember" facts about URLs, supported in "understanding" the significance of URLs' geological context, asked to "apply" context knowledge from the presentation, enabled to "analyze" results of numerical simulations, and invited to "evaluate" the effects of radioactive waste storage on the geology. For example, one task is the exploration of the tunnel system's position and orientation relative to the surrounding stratigraphic layers (see Figure 4). This motivates the users to visually combine information from two different spatial datasets to actively "understand" the geological context of the URL. Another task is to explore numerical simulation results of a determined in situ experiment that are available in the tour (see Graebling et al.<sup>3</sup> and the video<sup>a</sup> for details). In this case, users are asked to identify the time step when a simulated quantity reaches its maximum. In this way, they are encouraged to not only passively view the processes but to "analyze" the data to obtain a better understanding. The solution to all tasks is included in the prototype and can be accessed on demand by clicking a button.

Third, the application's user interface (UI) was improved by adding controls for the playback of audio comments, the display of text comments, and access to the tasks and their solutions (see top of

Figure 4). This supports participants in attending the tour at their own pace as all content can be accessed repeatedly and audio comments can be paused if necessary.

Fourth, digital assessments were set up so that the learning outcome can be tested while the users attend the tour. These tests are designed as inline tests (i.e., they follow the corresponding block of viewpoints immediately). The test infrastructure is implemented in a dynamic way, which means that tests are not hard-coded parts of the application, but can be provided as structured information in files of the JSON-format, which are then read by the application on runtime. This allows teaching experts to define their own tests based on their course's focus without changing the application itself. During the evaluation, the same tests were used for both courses.

### Heuristic Evaluation

In advance of the actual evaluation, a heuristic evaluation was conducted to identify and resolve major usability issues of the first prototype. An expert for user experience design analyzed the prototype and categorized their findings based on the heuristics by Nielsen.<sup>14</sup> In ten categories of usability matters, issues were assigned a severity rating from 1 (lowest severity) to 4 (highest severity). We resolved all issues with a severity of 2 or higher in a second version of the virtual tour prototype and discussed the adjustments with the user experience design expert. The full results of the heuristic evaluation are provided in the supplemental material that is available online.

### Test Evaluation

The evaluation procedure was tested before the actual evaluation events, using the second prototype. Three doctoral students from the Image and Signal Processing Group of Leipzig University participated in this test. The test evaluation revealed minor technical problems as well as the need for an option to invert the navigation controls. The technical issues were resolved and optional inverted controls were implemented ahead of the actual evaluation. By that, we created the final version of the virtual tour for the evaluation at the universities.

## STUDY DESIGN

The purpose of the evaluation is to answer two questions: 1) Are students from geoscientific courses of studies able to use the virtual tour in their courses without severe problems? and 2) How large is the

learning outcome when students attend the virtual field trip? The following sections present details on the study design.

## Target Group

The target group for the virtual field trip are graduate students from geoscientific courses of studies. Based on the process described by Gothelf and Seiden<sup>15</sup> and more recently by Jain et al.<sup>16</sup> proto-personas were created for this target group. This process is well established in the field of user experience research because it provides a fast and cost-efficient way to achieve a better understanding of the target group and its specific characteristics, motivations, and needs. For the process of the proto-persona creation, a group consisting of a visualization scientist, a user experience expert, and a scientist with teaching experience was formed. In preparation of the process, these three scientists studied demographics of geoscientific students and characteristics of geoscientific courses of studies. The group then brainstormed on relevant characteristics of the target group and created proto-personas based on these attributes. In a next step, similarities and differences of the created proto-personas were discussed and similar proto-personas were merged. This led to three final proto-personas that are provided in the supplemental material that is available online. The most important requirements derived from the final proto-personas were defined in a discussion by the group: 1) Participants should be able to attend at their own pace and repeated access to information should be available on demand. 2) The user interface should be simple and multilingual (German and English) and the information should be available as text and audio. 3) The participants should not be overburdened by interactions and information, but details should be available on demand.

## Measuring the Usability

### *Quantitative Evaluation*

The perceived usability was measured using UMUX-LITE,<sup>17</sup> which is a reduced version of the usability metric for user experience (UMUX). The questionnaire contains two items, that are answered on a 7-Point-Likert scale. Its shortness is helpful to avoid over-stressing the participants with too many questions. Lewis et al.<sup>18</sup> showed that UMUX-LITE scores can be used to calculate an approximated score for the system usability scale (SUS)<sup>19</sup> as “the correlation between the SUS and UMUX-LITE [is] significant and substantial.” Because SUS is an established questionnaire, it has been extensively analyzed. Its scores range from 0

(worst) to 100 (best) and can be categorized into more intuitive categories from D (worst) to A+ (best) or using the attributes “okay,” “good,” “excellent,” and “best imaginable” as described by Sauro and Lewis.<sup>20</sup>

### *Qualitative Evaluation*

In addition to the quantitative evaluation of the perceived usability, a short qualitative evaluation was performed as well. For this purpose, each item of the UMUX-LITE questionnaire was extended by asking the participants for an explanation why they gave that score. Furthermore, free-text inputs were added for suggested improvements and the participants’ impressions of the navigation and the scientific visualizations. The qualitative evaluation was conducted to not only measure a plain value that estimates the perceived usability, but also to identify possible problems and improvements.

## Measuring the Learning Outcome

All tests described below are multiple choice tests, in which exactly one of the three possible answers is correct. There was no time limit for completing the tests and it was possible to not answer a question if a participant was not sure of the correct answer. For reasons of feasibility, we did not use a larger variety of question types. This limitation could be addressed in future work.

### *General Knowledge*

As a first step, the general knowledge on the final disposal of radioactive waste with a focus on the situation in Germany was measured. This was achieved by comparing the results of a first test examining the prior knowledge before the virtual field trip with the results of a second test afterward. For this purpose, the full set of 20 questions was divided into two tests of 10 questions each. This was not done randomly to guarantee the same level of difficulty and to cover the full spectrum of content in both tests. The participants were randomly assigned to two groups of equal size. The two tests were presented to these groups in opposite order to remove potential bias, should the tests unexpectedly differ in complexity.

### *Specific Knowledge*

Besides the general knowledge, the virtual field trip presents specific knowledge concerning the Mont Terri URL. This covers: 1) information on the surrounding geological structure, the tunnel system, and an overview of the experiments, and 2) detailed information about the three selected experiments. The learning outcome for this specific knowledge was

measured by four tests: one on the URL and one per experiment. These tests are designed as inline tests (i.e., they follow the corresponding block of viewpoints immediately). Before each test, the participants are informed that there is no time limit for the test and that they have the possibility to go back to the viewpoints if they need more time for exploration. Because the contents of these inline tests are very specific, no prior knowledge was assumed and therefore there was no need for a prior test as conducted for the general knowledge.

### Hypotheses and Conditions

In order to avoid a biased interpretation of the results after the evaluation, hypotheses and conditions for their confirmation were defined in advance. All conditions concerning the learning outcome (H2 and H3) were defined in close collaboration with the teaching experts from the universities to match their expectations. Grading curves were defined in the same way. Since the results of the knowledge tests are processed anonymously and are not taken into consideration for the participants' grades, the rate of guessed answers is expected to be insignificant. In addition, participants were explicitly asked to avoid guessing answers. Hence, the hypotheses do not take "guessing" into account.

#### **Hypothesis H1**

The virtual tour's usability is perceived as at least good. H1 is considered true if the median SUS Score of the application is 71.1 or higher, which corresponds to the attribute "good."<sup>20</sup> As described in the "Measuring the Usability" section, the approximated SUS score is calculated from the UMUX-LITE items.

#### **Hypothesis H2**

The virtual field trip successfully transfers general knowledge on the final disposal of radioactive waste. H2 is considered to be true if the subhypotheses H2a (improvement) and H2b (good final grade) are true.

#### **Hypothesis H2a**

After the presentation, the general knowledge on the final disposal of radioactive waste is larger than before. H2a is considered to be true if for at least 90% of the participants (except for outliers) the number of correct answers increases by at least 20%. Ninety percent is a threshold that was defined in consultation with the teaching experts from the two universities as they normally experience failure rates of roughly 10% in courses of similar content.

#### **Hypothesis H2b**

After the presentation, the general knowledge on the final disposal of radioactive waste is at least "satisfactory." H2b is considered to be true if at least 90% of the participants (except for outliers) answer at least 55% of the questions on the general knowledge correctly and the median score is at least "good" (i.e., 70% or larger). The grading curve (as shown in Figure 7) was defined in consultation with the teaching experts from the two universities.

#### **Hypothesis H3**

The virtual tour successfully teaches special knowledge on the Mont Terri URL. H3 is considered to be true if the subhypotheses H3a (median result) and H3b (lower bound of individual results) are true.

#### **Hypothesis H3a**

The group of participants reaches good results in the special knowledge test. H3a is considered to be true if the median score is at least 70% (i.e., mark "good" or better).

#### **Hypothesis H3b**

Most of the students pass with at least satisfactory results. H3b is considered to be true if at least 90% of the participants (except for outliers) answer at least 55% of the questions on the special knowledge correctly (i.e., pass with mark "satisfactory" or better).

### Summary of the Evaluation Procedure

A short introduction and organizational details were followed by the technical setup. The participants then took an initial assessment, which measured their prior knowledge. Afterward, we gave a short introductory presentation about the final disposal of radioactive waste to provide context for the subsequent virtual tour. After this presentation, a second assessment was carried out to measure the learning outcome for this format. The participants then attended the actual virtual tour through the Mont Terri URL, including the four inline tests, at their individual speed. At the end of the tour, the participants filled out a form containing personal questions, their rating of the usability, and free-text feedback.

## EVALUATION RESULTS

### Participants and Hardware

#### **Participants**

The application was evaluated in two different courses at master level at two universities: in

**TABLE 1.** The participants' fields of study.

Location	Course of Studies	Participants
Freiberg	Geotechnics and Mining	8
	Geoinformatics	1
Dresden	Hydro Science and Engineering	11
	Groundwater and Global Change	2

"Numerical Methods in Geotechnical Engineering" at the TU Bergakademie Freiberg and in "Ground Water" at TU Dresden. The fields of study attended by the participants are listed in Table 1. Overall, 22 participants took part in the evaluation, 13 in Dresden and 9 in Freiberg. Fourteen of the participants identified as male and 8 as female. At both universities, four female participants attended. The age of the participants was in the range from 21 to 32 years with a median of 24.5 years. The average age in Dresden was approximately 27 years while the participants in Freiberg were much younger with an average of 22 years. One participant stated that they have a red-green deficiency. Of the 22 participants, 15 responded positively to the question of whether they enjoy playing 3-D games on a PC or console, which was asked to survey their gaming experience. None of the participants had detailed prior knowledge on the Mont Terri URL. Fourteen participants chose to attend the English version and eight the German version. Attending the virtual field trip as well as the evaluation was voluntary. The participants were informed before the event that their data are processed in anonymized form and that they can request the deletion of their data at any time. The results of the knowledge tests were not taken into account for their course marks.

### Hardware

At both universities the evaluation sessions were performed in computer laboratories, where the screen resolution was 1920 × 1080 pixels. We provided over-ear headphones to avoid influence and distractions while attending the tour. In Freiberg, the computers were equipped with an Intel Core i7-4790K CPU with 4 GHz and 32 GB of RAM. They did not have a dedicated GPU, but only an integrated Intel HD4600 and run Windows 10. In Dresden, the computers had an Intel i7-9700 CPU with 3 GHz and 32 GB of RAM. These computers were equipped with NVIDIA Quadro P400 graphic cards and run Windows 10.



**FIGURE 5.** Perceived Usability as SUS scores (0–100), calculated from UMUX-LITE answers, depicted with attributes and marks as described by Sauro and Lewis.<sup>20</sup>

## Usability

### Quantitative Results

Figure 5 shows a box plot of the virtual tour's perceived usability and the corresponding attributes and grades as defined in Sauro and Lewis.<sup>20</sup> Hypothesis H1 (good usability, see the "Hypotheses and Conditions" section) is verified as the median SUS score is 74.36. This value corresponds to the mark "B" and the attribute "good." The score that was given most often (seven times) is 77.07, close to the boundary between "B" and "B+" (both considered "good"). One participant rated the application's usability with a score of 49.98 which is worse than "okay." Their free-text input regarding problems and suggested improvements lacks clear, in-depth information about the reasons for this rating. The participant did not report a red-green deficiency. Two participants rated the application with a score of 87.9, which corresponds to mark "A+" and the attribute "best imaginable."

### Qualitative Results

The qualitative evaluation confirmed the findings from the quantitative evaluation of the perceived usability: None of the participants mentioned severe problems using the application. The minor problems described by the participants in the free-text input are consistent with their top three suggested improvements: 1) Participants wished for an improvement of the camera rotation interactions. The specific realization was not described in more detail by the participants. We assume that the possibility to rotate around objects in the scene instead of rotating only the camera around

its own axes might be an improvement. 2) The proposed smoother renderings of the visualizations presented in the application can be met by exclusively using computers with dedicated graphic cards when attending the virtual field trip. 3) Participants wished for different enhancements of the tasks inside the tour, like more feedback for their answer and even more details on the solutions.

*“FEELS LIKE AN INDIE GAME”*

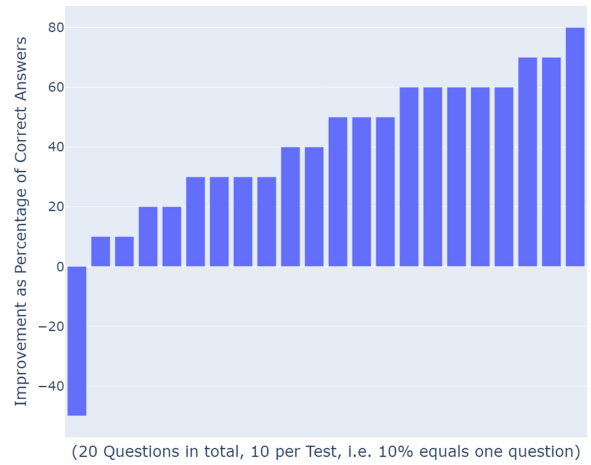
The most mentioned positive free-text feedback is the intuitive and interactive presentation of information, which provides vivid impressions of the URL while allowing to focus individual interests. Another frequently highlighted aspect is the advantage of being able to attend the virtual tour at your own speed. Participants appreciated the possibility to pause, continue, and repeat the display of information, as this supports the individuality of learning processes. One participant mentioned that the experience of the virtual tour “feels like an indie game,” which emphasizes that the application was perceived as a serious game. One of the professors was enthusiastic about the ability to use the tasks to direct students to specific aspects of the visualizations.

### Learning Outcome

#### General Knowledge

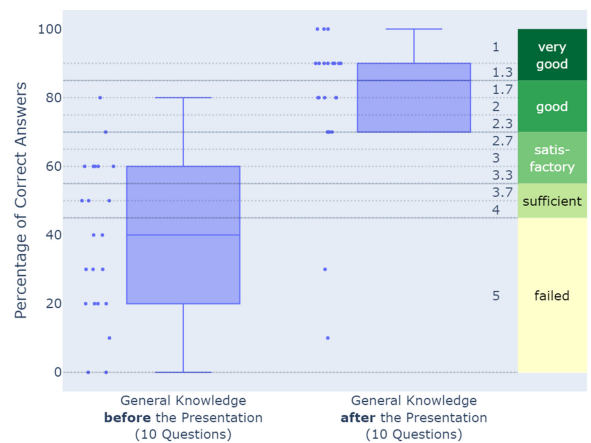
Figure 6 shows the individual improvement of the participants when they were tested on the general knowledge on the final disposal of radioactive waste before and after the presentation. The scores of all participants improved, with the exception of one outlier, who got 50% more questions wrong after the presentation than before. Therefore, we assume that they were not really on task and excluded their results from this analysis. Two participants only improved by 10%, which is lower than the condition formulated in hypothesis H2a (see the “Hypotheses and Conditions” section) and 19 participants improved by 20% and more, which is higher than the condition threshold. In addition, the median of correct answers increased from 40% to 85% (see Figure 7). Therefore, H2a is considered proven: The presentation significantly improved the general knowledge on the final disposal of radioactive waste.

Besides the pure improvement, it is also necessary to investigate the quality of knowledge after the presentation to make sure the participants not only improved, but also reached a sufficiently good level of knowledge



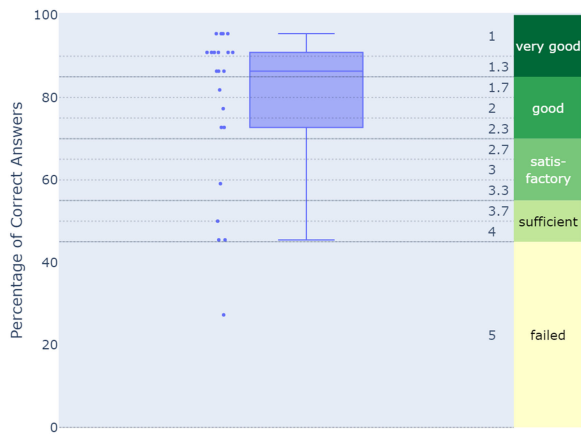
**FIGURE 6.** Individual improvement of the general knowledge on the final disposal of radioactive waste between prior and posterior test as difference of the percentage of correct answers.

(hypothesis H2a, see the “Hypotheses and Conditions” section). Figure 7 shows that after the presentation only two participants failed the general knowledge test, answering only 10% and 30% of questions correctly, respectively. The other 20 participants passed the test with the marks “good” or “very good,” which fulfills the criterion that at least 90% of the participants reach at least “satisfactory” results (55% or better). The median result of 85% correctly answered questions (on the border between “good” and “very good”) is also higher than the required 70%. Therefore, the conditions for hypothe-



**FIGURE 7.** Collective improvement of the general knowledge on the final disposal of radioactive waste between pre- and post-test in percentage of correct answers, with depiction of German university grades and corresponding adjectives.





**FIGURE 8.** Results of the 22 questions of the inline special knowledge tests per participant as percentage of correct answers.

sis H2b (see the “Hypotheses and Conditions” section) are clearly fulfilled.

As the hypotheses H2a and H2b were both verified, hypothesis H2 is considered true: The virtual field trip successfully transfers general knowledge on the final disposal of radioactive waste.

### Specific Knowledge

Figure 8 shows the results of the inline tests on the specific knowledge (i.e., on the Mont Terri URL and the three experiments focused on during the virtual tour). One participant failed the test because they answered only 27.27% of the questions correctly. This person is considered an outlier as their result is significantly worse than the box plot’s lower fence (approx. 45%) and is therefore removed from the analysis. The condition for hypothesis H3a is clearly fulfilled as the median score is 86.36%, which is significantly larger than the required 70% of correctly answered questions. Therefore, it is regarded verified that the group of participants reaches good results in the special knowledge test. Nonetheless, three participants only reached the mark “sufficient.” Although they passed the tests, their results are lower than the threshold for H3b. These three participants comprise circa 14% of the whole group. That means that the condition for H3b, that at least 90% of the participants answer at least 55% of the questions correctly, is narrowly missed. Still, it needs to be considered that the condition is only missed by one participant, accounting for approximately 4.7%. At the time of formulating the hypotheses and conditions, we planned with more participants. This would have led to a smaller effect of



**FIGURE 9.** Result comparison for three groups of questions sorted by how their answers were presented: in the presentation, in visualization and comment, and in visualization and task.

an individual participant’s score, so that participants who were not really on task would have had less influence on the overall result. Because of this unexpected situation and the excellent median results, it can be argued that the virtual tour successfully teaches special knowledge on the Mont Terri URL.

### Additional Findings

In this section, we will describe additional findings that do not correspond to the hypotheses described in the “Hypotheses and Conditions” section.

First, the scores varied among the four different special knowledge tests: On the one hand, the inline test on the Mont Terri URL and the test on one of the experiments (the full scale emplacement experiment, see Graebing et al.<sup>3</sup> for details on all experiments) showed exceptionally high median scores of 100% correct answers. On the other hand, the tests on the other two experiments resulted in lower median scores of 75% (cyclic deformation experiment) and 83.33% (fault slip experiment). Possible explanations are that the questions might have been inadvertently easier or that the participants missed some information during the interactive presentations. The free-text answers gave no clear indication on possible issues explaining this effect.

Second, the success of the knowledge transfer depending on the type of knowledge presentation within the virtual field trip has been investigated. The results are shown in Figure 9. For this purpose, we defined three types of knowledge transfer: knowledge from the conventional presentation (20 questions), information presented as visualization and comment (16 questions), and knowledge gained by working on

the corresponding task (6 questions). The results indicate that, in the scope of our study, the learning outcome is better when information is not presented in the conventional presentation (median of circa 61%) but in the interactive virtual tour as visualizations with comments (median of circa 77%) and tasks (median of 75%). However, this could also be biased by the different level of complexity of the subjects being presented in the different formats. Future work could compare the knowledge transfer efficiency of the different formats when presenting identical learning content.

Third, we compared the usability ratings concerning the participants' gaming experience. We assumed that the usability ratings of participants without gaming experience would be lower because they are not used to the relevant concepts like 3-D navigation. Therefore, it was important to test for differences and to make sure that the application is also usable without gaming experience. The median SUS score given by participants with gaming experience was 77.07 (mark B, "good") and the one given by participants without gaming experience was 71.65 (mark C+, "good"). The lowest score of a participant without gaming experience is 66.23, which corresponds to mark C ("okay") and all other participants without gaming experience rated the application's usability as "good." The results from both groups fulfill the condition of H1 (median score is 71.1 or higher). We conclude that the application has a good usability even for participants without gaming experience and that the influence of the participant's gaming experience on our application's usability is rather small.

Fourth, comparisons between the participants of the two different universities showed minor differences: Participants at Freiberg scored slightly better on average in the inline knowledge tests (Dresden: 74.82%, Freiberg: 82.82%). This can be explained by the fact that their geoscientific background is closer to the topic of the virtual field trip than the hydrological background of the participants in Dresden. The usability has been rated slightly better in Dresden than in Freiberg. However, the difference of the average SUS scores of 5.74 is quite small. Yet, a possible explanation is the better hardware setup in Dresden with computers equipped with dedicated graphic cards and therefore smoother renderings of the visualizations presented in the application.

## CONCLUSION AND FUTURE WORK

In this work, we adapted the virtual experiment information system of the Mont Terri URL for an

educational context. By extending the existing application and adding a conventional introductory presentation on the final disposal of radioactive waste, we created an advanced virtual field trip for intermediates (i.e., students), that can be used for university education. Further usage targeting other groups of intermediates, like the different actors of participatory processes during nuclear waste disposal site selection processes, is possible. A strength of our work is the interactive visualization of actual research data and its context, allowing intermediates to explore the data individually in an active learning process supported by the application's didactic design. The evaluation indicates a good perceived usability of the virtual tour prototype even for participants with low prior knowledge and little experience with 3-D applications. Furthermore, the evaluation gives evidence of a significant knowledge transfer provided by the virtual field trip. Some of the participants proposed to add even more in-depth information and asked detailed questions during the participation in the virtual field trip, which indicates that the application arouses curiosity. This also shows the importance of the on-site availability of an expert, which is a small limitation of the application's use-cases. Due to the good evaluation results, the two main questions of this work (see the "Study Design" section) can both be answered positively: 1) "Students from geoscientific courses of studies are able to use the virtual tour in their courses without severe problems." and 2) "The virtual field trip successfully provides a learning outcome that is appropriate for university education." In addition, in the scope of our work, interactive visualizations with audio comments and tasks showed a higher learning outcome than the conventional passive presentation. Overall, the evaluation clearly demonstrated the high potential of virtual field trips for the education of students.

Future work includes further adjustments of the existing application to explore the use for outreach and knowledge transfer to inform the public about the methodology and results of radioactive waste research. This could also be a starting point for discussions and exchange between researchers and the public, which could be valuable for participation programs within site selection processes. In addition, the virtual field trip prototype could be developed further to be actually used in academic teaching. Therefore, the assessments could be improved by including a larger variety of question types to adequately address the learning objectives and by providing immediate feedback to learners. In addition, improvements for the navigation controls for this special use case would need to be researched further. Beyond a focus on a

broader target group, future work could also focus more on single experiments instead of the overview of the whole URL. In this way, scientists could present the comprehensive life cycle of one experiment—from planning, through execution, to the data analysis—to the broad public. It is conceivable to implement such an experiment visualization as an interactive Virtual Reality experience to engage participants even more by using immersion.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Thomas Nagel, Dr. Andreas Hartmann, and Dr. Alireza Kavousi Heydari for the possibility to evaluate the authors' application in their courses and their support concerning educational and didactic decisions. They also thank the students of the two courses for their dedicated participation. Furthermore, the authors thank Dr. David Jaeggi, Dr. Gesa Ziefle, Dr. Dmitri Naumov, Dr. Wenqing Wang, swisstopo and Nagra for providing datasets and sharing their geological expertise. In addition, they thank Lars Bilke for his technical support of this work. Furthermore, the authors thank Yves Annanias, Michelle Philipp, and Christofer Meinecke for participating in the evaluation's pretest as well as Dr. Dirk Zeckzer for sharing his expertise on evaluations. Furthermore, the authors thank Susann Goldstein for proofreading the manuscript. Last but not least, the authors appreciate the support of the Mont Terri Project within the VR-A Task.

This work was supported by the European Joint Programme on Radioactive Waste Management EURAD under Grant 847593, in particular in the work package Monitoring equipment and Data Treatment for Safe repository operation and staged closure (MODATS).

## REFERENCES

1. P. Bossart and A. G. Milnes, *Mont Terri Rock Laboratory, 20 Years*. Cham, Switzerland: Birkhäuser, 2018.
2. M. R. Di Nucci and A. Brunnengräber, "The long road towards the soft nuclear repository state: Nuclear waste governance in Germany," in *The Future of Radioactive Waste Governance. Energy Policy and Climate Protection*. Wiesbaden, Germany: Springer VS, 2023.
3. N. Graebing et al., "Prototype of a virtual experiment information system for the Mont Terri underground research laboratory," *Front. Earth Sci.*, vol. 10, 2022, Art. no. 946627.
4. L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY, USA: Addison Wesley Longman, 2001.
5. J. Zhao, P. Lafemina, J. Carr, P. Sajjadi, J. O. Wallgrün, and A. Klippel, "Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based STEM education," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2020, pp. 893–902.
6. A. Klippel et al., "The value of being there: Toward a science of immersive virtual field trips," *Virtual Reality*, vol. 24, pp. 753–770, 2020.
7. L. S. Ferro, F. Sapio, A. Terracina, M. Temperini, and M. Mecella, "Gea2: A serious game for technology-enhanced learning in STEM," *IEEE Trans. Learn. Technol.*, vol. 14, no. 6, pp. 723–739, Dec. 2021.
8. D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the digital twin: A systematic literature review," *CIRP J. Manuf. Sci. Technol.*, vol. 29, pp. 36–52, 2020.
9. A. Mittal, L. Scholten, and Z. Kapelan, "A review of serious games for urban water management decisions: Current gaps and future research directions," *Water Res.*, vol. 215, 2022, Art. no. 118217.
10. Z. Liang, K. Zhou, and K. Gao, "Development of virtual reality serious game for underground rock-related hazards safety training," *IEEE Access*, vol. 7, pp. 118639–118649, 2019.
11. G. Dolphin, A. Dutchak, B. Karchewski, and J. Cooper, "Virtual field experiences in introductory geology: Addressing a capacity problem, but finding a pedagogical one," *J. Geosci. Educ.*, vol. 67, no. 2, pp. 114–130, 2019.
12. M. C. R. Harrington, Z. Bledsoe, C. Jones, J. Miller, and T. Pring, "Designing a virtual arboretum as an immersive, multimodal, interactive, data visualization virtual field trip," *Multimodal Technol. Interact.*, vol. 5, no. 18, pp. 1–25, 2021.
13. B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Taxonomy of Educational Objectives: Handbook I: Cognitive Domain*. New York, NY, USA: David McKay, 1956.
14. J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proc. ACM SIGCHI Conf. Hum. Factors Comput. Syst.*, Boston, MA, USA, 1994, pp. 152–158.
15. J. Gothelf and J. Seiden, *Lean UX: Applying Lean Principles to Improve User Experience*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2013.

16. P. Jain, S. Djasasbi, and J. Wyatt, "Creating value with proto-research persona development," in *HCI in Business, Government and Organizations. Information Systems and Analytics*. (HCII 2019. Lecture Notes in Computer Science), vol. 11589, Cham, Switzerland: Springer, 2019, pp. 72–82.
17. J. R. Lewis, B. S. Utesch, and D. E. Maher, "UMUX-LITE—when there's no time for the SUS," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, pp. 2099–2102.
18. J. R. Lewis, B. S. Utesch, and D. E. Maher, "Investigating the correspondence between UMUX-LITE and SUS scores," in *Design, User Experience, and Usability: Design Discourse* (Lecture Notes in Computer Science), Cham, Switzerland: Springer, 2015, pp. 204–211.
19. J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*. London, U.K.: Taylor & Francis, 1996, pp. 189–194.
20. J. Sauro and J. R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*, 2nd ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

**NICO GRAEBLING** is with the Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany. In addition, he is working toward the Ph.D. degree with the Image and Signal Processing Group of Leipzig University, Leipzig, Germany. His current research interest is interactive visualization in the geoscientific domain. Graebing received his master's degree from Leipzig University. He is the corresponding author of this article. Contact him at nico.graebling@ufz.de.

**MELANIE ALTHAUS** is a user experience designer focusing on functional and accessible designs, currently working with the Helmholtz-Centre for Environmental Research, 04318, Leipzig, Germany. Contact her at contact@melanie-althaus.de.

**ÖZGÜR OZAN ŞEN** is a researcher with the Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany. His research interests are immersive visualization of environmental data, virtual reality, game technologies, and computer graphics. Şen received his M.Sc. degree from the Hacettepe University, Ankara, Turkey. Contact him at oezguer-ozan.sen@ufz.de.

**THOMAS REIMANN** is a researcher and lecturer at TU Dresden, 01062, Dresden, Germany. He is also a guest lecturer for

hydrogeology at the University of Gothenburg, Gothenburg, Sweden. His research focuses on groundwater engineering, often in combination with distributed numerical models in different environments. Reimann received his Ph.D. degree from TU Dresden, Dresden, Germany. Contact him at thomas.reimann@tu-dresden.de.

**TUANNY CAJUHI** is a researcher with the Federal Institute for Geosciences and Natural Resources, 30655, Hanover, Germany. Her research interests include the study of coupled thermo-hydro-mechanical phenomena and fracture in porous media. Cajuhi received her Ph.D. degree in the field of numerical methods from the University of Technology, Braunschweig, Germany. Contact her at tuanny.cajuhi@bgr.de.

**GERIK SCHEUERMANN** is a professor with Leipzig University, 04109, Leipzig, Germany, and holds the chair of the Image and Signal Processing Group. His research interests include topology, Clifford algebra, image processing, information visualization, and scientific visualization. Scheuermann received his Ph.D. degree in computer science from the University of Kaiserslautern, Kaiserslautern, Germany. He is a member of IEEE. Contact him at scheuermann@informatik.uni-leipzig.de.

**OLAF KOLDITZ** is the head of the Department of Environmental Informatics, Helmholtz Center for Environmental Research, 04109, Leipzig, Germany, and he is also with the Dresden University of Technology, Dresden, Germany. His research interests include environmental fluid mechanics, numerical methods and software engineering with applications in geotechnics, hydrology and energy storage. Kolditz received his Ph.D. degree in natural sciences from the Academy of Science of the GDR and earned his habilitation in engineering sciences from Hannover University. Contact him at olaf.kolditz@ufz.de.

**KARSTEN RINK** is a postdoctoral fellow with the Helmholtz Centre for Environmental Research, 04318, Leipzig, Germany, and a group leader for data integration and visualization. His research interests include the integration of heterogeneous geoscientific data collections as well as the immersive visualization of environmental data. Rink received his Ph.D. degree in visual computing from the University of Magdeburg, Germany. Contact him at karsten.rink@ufz.de.