Computers & Graphics (2024)

Contents lists available at ScienceDirect

Computers & Graphics

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

Enhancing Anatomy Learning Through Collaborative VR? An Advanced Investigation

Haya Almaree[a,*], Roland Fischer[a], René Weller[a], Verena Uslar[b], Dirk Weyhe[b], Gabriel Zachmann[a]

[a] *University of Bremen, Bremen, 28359, Germany*
[b] *University Hospital for Visceral Surgery, University of Oldenburg, PIUS-Hospital, Oldenburg, 26121, Germany*

ARTICLE INFO

Article history:
Received July 25, 2024

Keywords: Virtual Reality, Anatomy, Education, Collaborative Learning

ABSTRACT

Common techniques for anatomy education in medicine include lectures and cadaver dissection, as well as the use of replicas. However, recent advances in virtual reality (VR) technology have led to the development of specialized VR tools for teaching, training, and other purposes. The use of VR technology has the potential to greatly enhance the learning experience for students. These tools offer highly interactive and engaging learning environments that allow students to inspect and interact with virtual 3D anatomical structures repeatedly, intuitively, and immersively. Additionally, multi-user VR environments can facilitate collaborative learning, which has the potential to enhance the learning experience even further. However, the effectiveness of collaborative learning in VR has not been adequately explored. Therefore, we conducted two user studies, each with $n_{1,2} = 33$ participants, to evaluate the effectiveness of virtual collaboration in the context of anatomy learning, and compared it to individual learning. For our two studies, we developed a multi-user VR anatomy learning application using UE4. Our results demonstrate that our VR Anatomy Atlas offers an engaging and effective learning experience for anatomy, both individually and collaboratively. However, we did not find any significant advantages of collaborative learning in terms of learning effectiveness or motivation, despite the multi-user group spending more time in the learning environment. In fact, motivation tended to be slightly lower. Although the usability was rather high for the single-user condition, it tended to be lower for the multi-user group in one of the two studies, which may have had a slightly negative effect. However, in the second study, the usability scores were similarly high for both groups. The absence of advantages for collaborative learning may be due to the more complex environment and higher cognitive load. In consequence, more research into collaborative VR learning is needed to determine the relevant factors promoting collaborative learning in VR and the settings in which individual or collaborative learning in VR is more effective, respectively.

1. Introduction

The teaching of human anatomy is fundamental in medical education as it forms the basis for the development of clinical and surgical knowledge among professionals[1, 2], and influences the design of the medical curriculum [3]. Classi-

*Corresponding author:
e-mail: haya@uni-bremen.de (Haya Almaree)

cally, anatomy teaching is done using dissection, prosection, anatomical models, and lectures. Dissection offers a hands-on approach to examining anatomical specimens, enhancing students' understanding of anatomy [4], prosection reinforces students' comprehension of complex structures and relationships, and anatomical 3D models help to visualize anatomical structures. However, dissection is costly and time-consuming, prosection relies heavily on the anatomist's skill and expertise [5], and lectures may not be effective in promoting active learning and engagement compared to more interactive approaches. Also, the availability of human cadavers and animal specimens for dissection is limited [6].

Thus, virtual reality (VR) has become increasingly prevalent in recent years and is considered a valuable tool in education [7, 8]. The technology offers several advantages, such as safe, controllable, immersive 3D environments and natural interaction, making the learning experience more intuitive and engaging [9, 10]. However, most current VR-based learning applications are limited to single-user usage, and there is minimal research on the effectiveness of collaborative VR-based learning. Collaborative learning, in general, has been shown to have positive effects on learning outcomes [11, 12, 13, 14] and to provide numerous other benefits, though. For instance, a higher problem-solving performance, a shared understanding of meanings and a shared sense of achievement [15], increased productivity, positive interpersonal relationships [16], better psychological health, higher social competence, and self-esteem [17]. In a competitive anatomy learning setting, Du et al. [18] found that the multi-user VR group had higher learning outcomes than the single-user VR one, but also significantly higher stress levels.

To investigate whether collaboration in VR, in our example for anatomy learning, offers benefits and more positive learning results compared to individual VR learning, we developed a VR anatomy learning application that is multi-user capable. In order to evaluate its effectiveness, we conducted two user studies, the latter introducing an additional knowledge-consolidating phase. Concretely, we examined the participants' learning progress, usability, and motivation when using our VR learning application, both individually and in groups. We also explored potential relationships between these measures and demographic data. With this work and our findings, we offer valuable insights into this under-researched topic. This paper extends our previous publication [19] with an extended evaluation and a second user study. However, all the content and results of our previous publication are included in this extended version, too.

2. Related Work

Virtual reality (VR) is a rapidly expanding field that holds promise for a variety of applications in healthcare, most importantly for education and training. Accordingly, the use of VR in medicine got much attention lately [20]. For example, Falah et al. [21] developed a VR and 3D visualization system for anatomy teaching that offers an interactive, real-time 3D representation of the human heart and various self-assessment tools. Similarly, Fairen et al. [22] developed and evaluated a VR anatomy teaching tool that provides real-time, interactive 3D representations of various anatomical structures that were augmented with additional information. An evaluation with anatomy students showed very positive results. Codd and Choudhury [23] evaluated the use of 3D virtual reality and compared it with traditional anatomy teaching methods (dissection and textbooks) on the example of a human forearm. Interestingly, they found no significant learning advantages using VR. In contrast, Kurul et al. [24] also conducted a study on anatomy training comparing immersive, interactive 3D VR with classical teaching methods and found the former to lead to significantly higher test scores. Another example highlighting the benefits of VR to anatomical education is the Immersive 3D Anatomy Atlas by Gloy et al. [25]. It provides a realistic 3D model of the human body in an immersive environment and allows users to explore individual anatomical structures interactively. An evaluation showed that the VR group took significantly less time to answer anatomical questions and had significantly better test results than students who learned using textbooks. The use of VR in the medical area is not limited to education, though. Other promising application domains are surgery planning and training. For instance, Reinschluessel et al. [26] developed a VR-based surgery planning tool that provides a 3D view of medical data. They found that planning in VR had many advantages, such as improving the surgeons' spatial understanding and identifying anatomical structures.

Most VR tools for medical education and training are limited to single-user usage only, though. Only a few works enable collaboration and even fewer investigate its effects and benefits. Works that do provide shared VR environments are, for example, the one by Kaluschke et al. [27], who presented a multi-user haptic VR system for dental surgical skill training, and the one by Fischer et al. [28], who presented a system for real-time volumetric medical image visualization with support for multi-user VR interactions. Furthermore, Zorzal et al. [29] developed a collaborative Augmented Reality (AR) tool, Anatomy Studio, for virtual dissection, that allows to create and manipulate 3D reconstructions from cryosection images using tablets, AR headsets, and mid-air interactions. A user study with medical professionals showed positive results regarding enhanced discussion and spatial understanding, making it a valuable tool for anatomical education and training. Later, Jorge et al. [30] introduced Anatomy Studio II, which features a database of 3D meshes and a client-server-based distributed architecture for collaborative VR & AR remote education and training. The focus, however, is on virtual dissection. This involves the reconstruction of individual 3D anatomical structures based on contouring them in 2D slices. Recently, Wang et al. [31] investigated the integration of VR-based cooperative learning strategies within a gross anatomy curriculum and showed that the collaborative VR group had better learning outcomes compared to the group that learned through classical methods. Similarly, Du et al. [18] developed an anatomy learning game and found the VR groups to have higher learning outcomes than the textbook group. Interestingly, they also compared single-user VR and multi-user VR and found the latter to have higher learning outcomes but also significantly higher stress levels. The game was competitive and not collaborative by nature, though. Boedecker et al. [32] also developed an immersive VR application for liver surgical planning that was later extended by Schott et al. [33] to allow for collaborative usage. It provides various teaching scenarios for collaborative and cooperative training in different group sizes. An exploratory study with medical students and surgery lecturers indicated positive outcomes for usability and presence. Another immersive VR learning environment that supports collaboration of multiple users was developed by De Back et al. [34, 35]. Its effectiveness was shown through two empirical studies that revealed that collaborative learning provided greater learning gains compared to conventional textbook learning, particularly among participants with low spatial ability. In contrast, Sedlák et al. [36] compared collaborative and individual learning of geography in VR and while both groups learned effectively, there were no significant differences between groups for learning gain, speed, or motivaton. For a more detailed overview and review of VR for anatomy education, we refer to the works by Lee et al. [37] and van der Meer et al. [38].

3. Our Collaborative VR Anatomy Atlas

For our work, we decided to use the Immersive Anatomy Atlas by Gloy et al. [25] as a basis. It already provided a good implementation of a VR anatomy learning application and was based on the modern Unreal Engine 4. The latter made it easy to extend the application for our purposes, mainly, multi-user functionality. Thus, our Collaborative VR Anatomy Atlas allows multiple users to meet, interact, and collaboratively learn within a shared environment. An overview of the virtual environment, including the human anatomy model, a table with medical instruments, a virtual tablet, and a monitor, can be seen in Figure 1 (left), while Figure 1 (right) depicts two users exploring the anatomy model during a typical learning session. In addition to the multi-user functionality, we implemented a model of the human circulatory system, extensive logging functionality, multiple custom levels, and a VR quiz. We also made minor improvements to the meshes, textures, and tools, upgraded to the latest version of the Unreal Engine 4, and fixed a few bugs. The rest remained as in Gloy's original implementation.

The user interface is the same for the single- and multi-user conditions and consists of the HMD and controllers for interaction and room-scale and teleportation for locomotion. Each user is represented by an avatar consisting of a virtual HMD and a pair of hands with which they are able to grab, move, rotate, and interact with the organs, see figure 1 (right). When an organ is grabbed, it gets highlighted in white and its name is shown on a label. When a grabbed organ is held near the correct (original) position, the latter gets highlighted in green (depicted in Fig. 2 (right)). The Anatomy Atlas (as the original implementation) also provides an array of medical instruments, such as a cutout tool for hiding local geometry (Fig. 2 (second from right) or a pair of markers. When an organ or instrument is actively being held by a user, the other user is blocked from grabbing it, too. We decided on this approach in order to prevent users from deliberately or accidentally taking away objects from others and to prevent synchronization issues/conflicting control. Additionally, we chose this avatar model over more sophisticated (personalized) full-body avatars because it doesn't require complicated scanning/capturing setups, is computationally cheap, and is not prone to distracting or glitchy behavior (odd/wrong angles of untracked joints) that is common with inverse kinematics-based full-body avatars.

We use a client-server model based on the network functionality provided by the Unreal Engine 4, which allows for shared learning sessions between users in the same local network or over the internet. Technically, an arbitrary number of users can join but we focused on sessions with one or two users. The choice to employ the Unreal Engine helped greatly with implementing the multi-user functionality, as the engine already provides all the necessary components making it a straightforward process. Since we did not want to use or connect to any online and cloud-based services (e.g. Steam), we used the prototyping/testing variant of Unreal Engine's built-in onlinesubsystem [39] in our implementation, which still allows for finding and joining sessions in the local network or online via IP. Our application can be started as either a listen or a dedicated server. The avatars, body parts, and other interactive objects, such as the operation table, instruments, and tablets, get replicated (synchronized) between users using remote procedure calls (RPCs). Specifically, when an object is moved (significantly) or its state is changed by a user (client), an RPC is sent to the server to replicate the action/change there. Then, the server executes a multi-cast RPC to all connected clients to replicate the action/change there too. In our implementation, we prevent updates on the client that initiated the change, as the respective action was already performed locally. This improves responsiveness and reduces network traffic. We decided on this RPC-based approach, as we found it to give us more control over what exactly is synchronized, when, and how often. We experienced high network traffic and synchronization issues with a first, simple implementation and when using the standard settings on networking and replication. However, low latencies and reasonable bandwidth usage are important factors for multi-user (VR) systems and a smooth workflow [40]. Hence, we optimized our implementation and the replication process by employing struct replication (pooling multiple variables to be replicated in structures), delta replication (sending the numerically smaller difference in time-varying values than the absolute ones), caching, and careful selection of reliable/unreliable replication channels, reducing the data to be transmitted to a minimum.

We also developed and integrated additional features such as a model of the human circulatory system that simulates pulsatile blood flow. Using dynamic, animated textures, we illustrate the functionality of veins transporting blood toward the heart and arteries directing it away (see Fig. 2 (2 left-most images)). Furthermore, we implemented extensive logging functionality to enable researchers to track user interactions and behavior within the virtual environment. In order to make use of new features and improvements, we decided to upgrade/port the software to

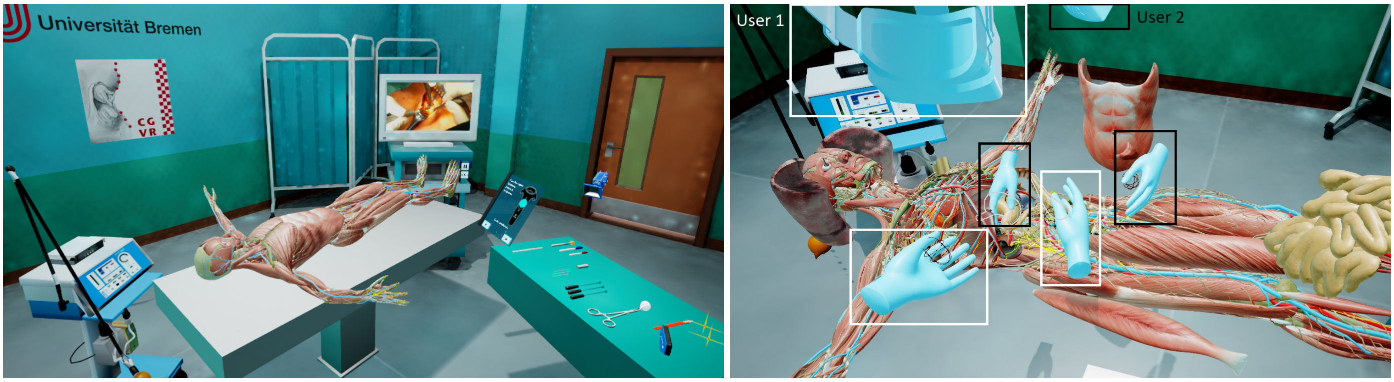

Fig. 1: Left: An overview of our Anatomy Atlas and its virtual environment. The human anatomy can be seen in the center, while the virtual tablet and the table with the medical instruments are on the right. Right: Two users within our Collaborative VR Anatomy Atlas, examining a human anatomy. Each user has a light-blue colored avatar consisting of a virtual head-mounted display (HMD) and a pair of hands. The avatar of user 1 is highlighted with white boxes and user 2's avatar with black boxes.

Unreal Engine version 4.27.X, although this meant additional work and bug fixes.

Lastly, a VR quiz (post-test) was developed to assess participants' anatomy knowledge following their learning session (in Study 1). Access to the quiz, along with other functionalities and controls, is facilitated through a button on an interactive tablet within the VR environment. For the second study, we instead created several levels specifically designed for different phases and tasks, e.g. no highlighting of correct organ positions in the test level. We used the implementation described in this section for both of our user studies, with no functional changes in between.

4. Research Questions

In collaboration with medical doctors and based on their medical opinions, we designed two user studies to investigate the effectiveness of collaborative anatomy learning in VR, specifically, using our collaborative VR Anatomy Atlas, and to compare it to individual learning. Moreover, we are also interested in assessing its impact on learning motivation and its usability. The studies are mostly aimed at novice medical students who have had no proper anatomy education yet. We formulated the following research question that we intend to answer with our studies: (R_1) Is the Collaborative VR Anatomy Atlas effective for anatomy learning? With this we want to make sure that our application and implementation are generally effective for anatomy learning, also, these research questions serve as a prerequisite for the following ones. Based on prior research that found benefits in collaborative learning [17, 16], we want to investigate if (R_2) collaborative learning in VR also leads to better learning outcomes than individual VR learning. Additionally, we want to evaluate the usability (R_3) and user experience/motivation (R_4), in general, and especially if there are any differences between individual and collaborative learning.

5. User Study 1

5.1. Design and Setup

For our first study, we employed a between-subject design, hence, we divided the participants randomly into two groups: one group testing the single-user condition and the other group testing the multi-user condition. Multi-user sessions always consisted of two participants and each condition was performed an equal number of times, thus, the number of participants testing the multi-user learning condition was twice as large. We limited ourselves to groups of two, in order to still get meaningful results while having a manageable sample size. The study was conducted in our laboratory and, in the case of the multi-user condition, both participants were in the same room and could communicate verbally, (see Fig. 3 (left) for an example).

The learning sessions using our Collaborative VR Anatomy Atlas were conducted using HTC VIVE Pro HMDs including a pair of controllers. To provide a good user experience, we ensured that the frame rate was maintained at 90 frames per second. In the virtual environment, the participants were represented through avatars (see Chapter 3) and were able to freely move around using room-scale VR and teleportation. The virtual environment resembled an operating room and included a virtual anatomic 3D model that they were supposed to interact with and explore in order to learn about the anatomy. An overview of the virtual environment is given in Figure 1 (left).

To evaluate the learning effectiveness, we designed a multiple-choice test consisting of 8 anatomy questions. The test, as well as the learning session itself, were carefully designed together with medical doctors to be manageable by novice medical students and even ordinary students. This test was conducted two times: one time before the learning session on paper (pre-test), and one time after the learning session directly in VR (post-test). For the latter, the participants transitioned to a quiz level. There, the correct answer is displayed in green, and the incorrect ones in red. The key presses for each answer were logged, but only the first answer entered was evaluated. Thus, the participants could learn the correct answer and improve their knowledge without affecting the validity of the study, even if they initially answered incorrectly. By comparing

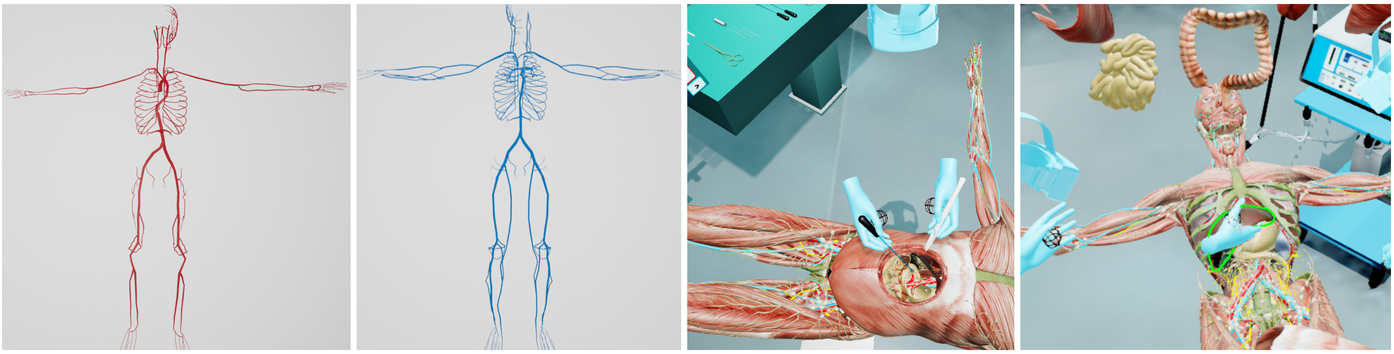

Fig. 2: The two left images illustrate the animated 3D model of the circulatory system: the left one depicts the arteries (red) and the right one the veins (blue). The two images on the right depict other specific features of our Anatomy Atlas: the left one shows a user using a cutout tool that hides geometry in a local area and the right one shows the in-application highlighting of the organ's correct position in green (when near to it).

the results of the two tests, we calculate the learning progress. Additionally, we employed questionnaires on usability and motivation. Specifically, the System Usability Scale [41] and an adapted version of the questionnaire on motivation for cooperative and playful learning strategies (CMELAC) [42]. We customized the latter by removing the "Teamwork" factor as it was not applicable in the single-user condition and we wanted to ensure equivalence between both conditions. However, we believe it to be still valid and reliable. We also added a question to gauge the participants' interest in learning in a virtual reality environment. To analyze the participants' behavior, we tracked the time they spent in VR, video recorded the sessions, and employed extensive data logging using our custom implementation.

5.2. Procedure

The procedure of our first study is depicted in Fig. 3. First, the participants were informed about the study and its goal, read and signed a consent form, and had time to ask questions. Then, the participants were asked to complete a demographical questionnaire about age, gender, previous experience with VR, etc. To determine the anatomical pre-knowledge, the participants were then asked to complete our pre-test questionnaire consisting of 8 anatomical questions (on paper). Following this, the Collaborative VR Anatomy Atlas application, its features, and its usage were briefly explained. Lastly, the participants were given up to three minutes to freely explore the VR environment and familiarize themselves with it.

Once the participants were ready, the learning session was started in which they had to explore the virtual anatomic model and complete various tasks with which we aimed to mimic classical non-VR learning. Specifically, the tasks were discovering the human anatomy, searching for specific organs (e.g. the spleen, pancreas, liver), and finding answers to the pre-test questions. The tasks were solvable individually as well as teamwise (in the multi-user condition), however, we expect the latter to be more effective, as in traditional learning. No assistance was given during task completion, but the tasks were repeatable. Figure 3 (left) shows an example of a multi-user learning session. Participants were given an unlimited amount of time.

Upon completion of the tasks, the participants were transitioned to the quiz level and took our anatomy post-test. There, they had to answer the shown questions by pressing the corresponding 3D buttons. After the post-test, and while their memories were still fresh, the participants had to complete the questionnaires about usability and motivation (on paper). They were also asked if they experienced any motion sickness and to provide subjective feedback. The procedure was identical for both conditions, with the exception that the participants of the multi-user group were explicitly instructed to work together on the anatomical tasks and to learn collaboratively. However, at the VR quiz level, they were required to complete the post-test independently.

5.3. Results

In this section, we present the results from our first study, including demographic data, the results of the anatomy pre-test and post-test as well as the results of the questionnaires on motivation and usability. As the data was, as expected, normally distributed, we conducted independent samples t-tests to test for statistically significant differences between the single and multi-user groups and employed a threshold of $p = 0.05$ (as usually done).

5.3.1. Demography

The study was conducted with $n = 33$ participants who were randomly divided into two groups for the two different learning modalities: 11 participants experienced solo learning (single user) and 22 participants experienced shared learning (multi-user) in random pairs. We selected participants who were roughly in the same age group as typical medical students. As they were mostly university students from various subjects, they had no particular medical experience. However, we assume the anatomical knowledge level to be roughly similar to novice medical students. The single-user group was made up of 2 female (18.2 %) participants and 9 male (81.8 %) participants, see Fig. 4 (left), while the multi-user group was made up of 14 men (63.6 %) and 8 women (36.4 %), see Fig. 4 (right).

Moreover, a substantial percentage of single users (54.5 %) and a smaller percentage of multi-users (22.7 %) reported having extensive experience with VR, a significant percentage of single users (36.4 %) and multi-users (31.8 %) reported having used VR before, while a minority of single users (9.1 %) and

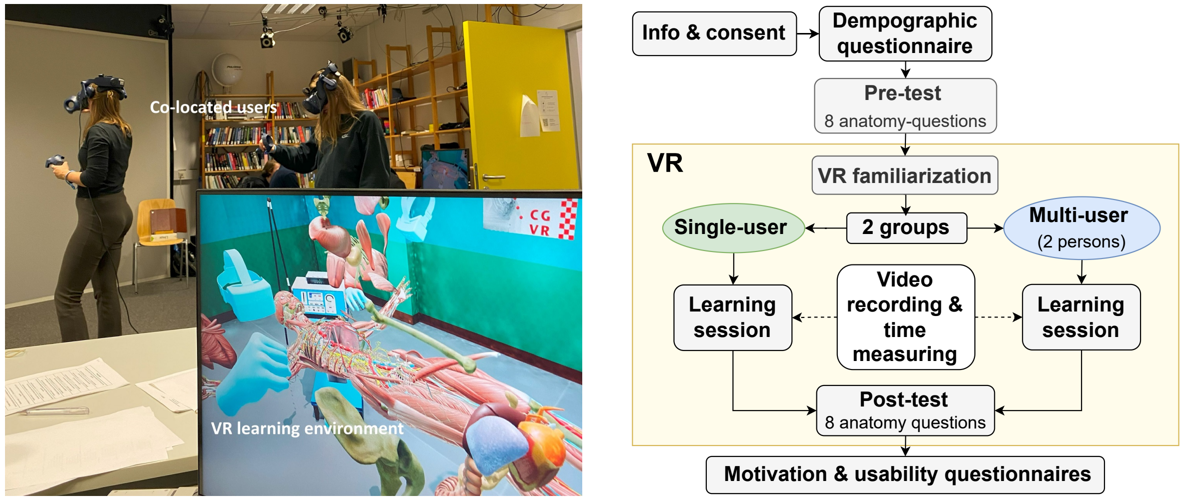

Fig. 3: Left: Multi-user learning session. In this case, both participants were in the same physical room. The virtual operating room with the anatomical model can be seen on the monitor in the foreground. Right: Flow chart of the study procedure. After the anatomy pre-test, participants performed the VR learning session in the virtual operating room, either alone or as a group, took the anatomy post-test individually, and eventually completed additional questionnaires on paper.

a substantial percentage of multi-users (45.5 %) had no experience with VR, see Fig. 4. Asked about the preferred learning setting, 45.5 % of the single users stated to generally prefer learning alone and only 9.1 % would prefer learning in a group. For the multi-user group, the ratio was more balanced with 27.3 % each, see Fig. 4. Lastly, regarding learning types, a majority of single users (63.6 %) and half of the multi-users (50.0 %) identified themselves as a visual learner. In contrast, a minority of single users (27.2 %) and multi-users (18.2 %) reported to be the intellectual learning type, and no single users stated to be the auditory learning type. Interestingly, a substantial percentage of multi-users (22.7 %) identified themselves as the auditory learning type, while a minority of both single users and multi-users (9.1 %) identified themselves as the haptic learning type, see Fig. 4.

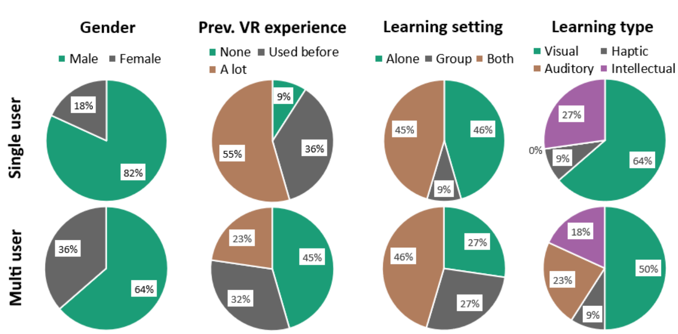

Fig. 4: Demographic results. Top row: single-user results; bottom row: multi-user results. From left to right: gender distribution, previous experience with VR, preferred learning setting, self-reported learning type. Most participants were men, especially in the single-user group. They also had more VR experience, on average. More participants in the multi-user group preferred to learn collaboratively and identified as auditory learners, relatively speaking.

5.3.2. Anatomy Knowledge, Learning Progress, and Time

The results of the anatomy pre-test with 8 anatomical questions (conducted before the VR learning session) are depicted in Fig. 5 (left). The mean pre-test score for the single-user group

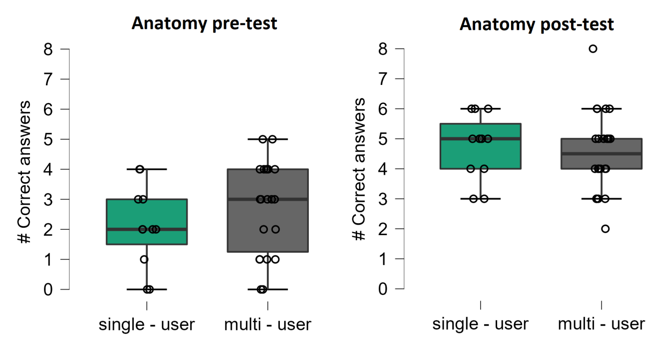

Fig. 5: Results of the anatomical knowledge pre-test conducted before the study (left) and the post-test conducted after the study (right). The multi-user participants had, on average, slightly more pre-knowledge. In the post-test, both groups scored better than before and fairly similar, with a slight advantage for the single-user group.

was 2.091 ($SD = 1.375$) and the one for the multi-user group was 2.727 ($SD = 1.518$). Although the means are similar, there is a slight advantage for the multi-user group. However, the difference is not statistically significant ($t(31) = -1.170, p = 0.251$).

The results of the anatomy knowledge post-test (conducted after the study) are depicted in Fig. 5 (right). Both groups visibly improved compared to the pre-test and answered more questions correctly. Between the groups, the results are again similar, this time, with just a slight advantage for the single-user group. The mean score for the single-user group was 4.727 ($SD = 1.104$) and the mean score for the multi-user group was 4.545 ($SD = 1.371$). We, again found no significant differences between the groups ($t(31) = 0.381, p = 0.705$).

In order to better investigate the learning effectiveness, we compute the participants' learning progress as the difference (delta) between the pre- and post-test results, see Fig. 6 (left). The single-user group, on average, did have slightly higher learning progress: the mean score was 2.636 ($SD = 1.859$), whereas the multi-user group's mean score was 1.818 ($SD =$

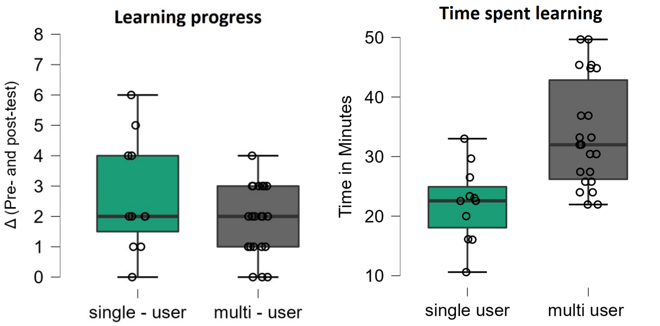

Fig. 6: Left: Learning progress (delta between pre- and post-test) for single- and multi-user groups. The single-user group learned, on average, slightly better. Right: Time spent learning in VR. The multi-user group spent, on average, more time in VR.

1.140). The median, however, is more similar between the groups. A t-test resulted in: ($t(31) = 1.569, p = 0.127$). However, the result is still above the usual threshold of $p \leq 0.05$ for statistical significance.

The time spent in the VR learning session, divided by single- and multi-user group, is depicted in Fig. 6 (right). The mean time for the single-user group was 22.130 minutes, whereas it was 33.774 minutes for the multi-user group. We found that the single-user group spent significantly less time in the VR environment than the multi-user group ($t = -3.783, p < 0.001$).

5.3.3. Questionnaires on Motivation and Usability

The results of the questionnaire on motivation for cooperative and playful learning strategies (measuring the factors of motivation, learning, and flow) are depicted in Fig. 7 (left). The average scores between the single- and multi-user groups are similar and both very positive. Concretely, on motivation, the means scores were 4.429 ($SD = 0.564$) (single user) and 4.253 ($SD = 0.612$) (multi-user), on learning 4.091 ($SD = 0.628$) (single user) and 4.164 ($SD = 0.564$) (multi-user), and on flow 3.97 ($SD = 0.69$) (single user) and 3.788 ($SD = 0.739$) (multi-user). The standard deviations indicate that the scores were relatively consistent within each group. We found no significant differences between the single-user and multi-user groups in terms of motivation ($t(31) = 0.795, p = 0.433$), learning ($t(31) = -0.336, p = 0.739$), or flow ($t(31) = 0.681, p = 0.501$).

The perceived usability of the Collaborative VR Anatomy Atlas was measured using the System Usability Scale. The SUS scores were calculated using the standard methodology and are depicted in Fig. 7 (right). Overall, the participants provided positive feedback and moderate to high ratings. The mean SUS score for the single-user group was 75.227 ($SD = 8.976$) and for the multi-user group 66.364 ($SD = 14.15$). The t-test revealed that there is a noticeable difference in means between the single-user and multi-user groups, although the usual threshold of $p = 0.005$ for statistical significance was just not reached ($t(31) = 1.887, p = 0.069$).

5.3.4. Correlations

To examine the relationships between measures and test for correlations, we calculated the Pearson correlation coefficient between various measures. The results are shown in Table 1. Between most measures, we found no or only very weak correlations. However, between age and motivation, there seems to be a strong positive correlation, and between gender and motivation a moderate negative correlation (higher motivation for women). However, although noticeable, none of these correlations reached the threshold for statistical significance.

Measure1	Measure2	Coeff. (r)	P-val.	Interpr.	Sig.
Learn. Prog.	SUS Score	0.1671	0.353	(+)	No
	Time	0.0365	0.840	o	No
	Age	0.0146	0.936	o	No
	Gender	0.0042	0.981	o	No
SUS Score	Age	-0.197	0.273	(-)	No
	Gender	0.029	0.874	o	No
Motivation	Age	0.778	0.121	++	No
	Gender	-0.645	0.239	-	No

Table 1: Overview of our correlation analysis. Only between motivation and age/gender seems to be a notable or strong correlation.

6. User Study 2

To further investigate the potential of collaborative anatomy learning, we reflected on and revisited the design and limitations of Study 1 and, again in collaboration with medical doctors and based on their professional opinions, designed a second user study. In this study, we focused on changing/adapting the task, specifically, we included a practical consolidation phase in which the participants had to (collaboratively) apply their newly learned knowledge, thereby consolidating it. To accommodate the new consolidation phase, we adapted the study design and procedure.

6.1. Design and Setup

We again employed the between-subject design with single-user and multi-user (pairs of two) conditions as in Study 1. In this second study, we added a new consolidation phase in which the participants had to apply their newly acquired knowledge by engaging in a (collaborative) practical, hands-on activity within the VR environment (correctly placing organs). An example illustration of this can be seen in Figure 8 (left). This phase aimed to help consolidate the learned knowledge through active participation and foster potential benefits from collaboration, thereby going beyond the sole memorization task from the learning phase (which might not benefit so much from collaboration).

We also changed the pre- and post-tests to follow a more hands-on, visual approach that should be closer to reality and the actual medical practice, and thus be possibly more effective. Specifically, both instances of the test were conducted in VR and participants had to find anatomically misplaced or superfluous organs in the 3D Anatomy Atlas, identify their names, and determine their correct locations (earning one point each to a total of 28). Figure 8 (right) depicts the test scene, for clarity, a couple of misplaced organs are highlighted in white. The learning session itself was principally conducted as before, but now the time was limited and the study assistant more strictly

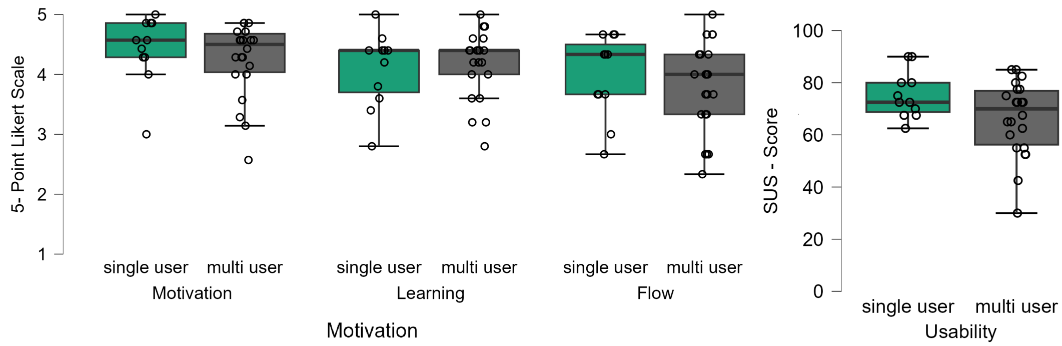

Fig. 7: Left: Results of the motivation questionnaire (factored into motivation, learning, and flow). The averages between the single- and multi-user groups are similar and both very high. Right: The System Usability Scale scores. The scores are generally high but the single-user group's feedback is more positive.

enforced a focused workflow. Generally, we again carefully designed the study such it is easy enough to be manageable by ordinary students who we assume to have similar knowledge as novice medical students. For all four main VR phases, we created different levels/rooms that were configured accordingly: learning (including medical instruments), consolidating (organs on the table), pre- and post-test (misplaced organs). Also, we again employed the questionnaires on usability and motivation, tracked the time the participants spent for each phase, and did extensive data logging and video recordings of the sessions.

For the study, we used HTC VIVE Pro HMDs and controllers, ensured to have always 90 fps, employed head+hands avatars (see for example Figure 8 (left)) for all participants, and allowed for free movement in the virtual operating room environment using room-scale VR and teleportation. As we focused on adapting the task and procedure for Study 2, which we found to be the most promising, the application itself remained functionally the same. With this approach of only changing one factor at a time we also avoid ambiguity in what caused potentially different results.

6.2. *Procedure*

The procedure of our second study, depicted in Fig. 9, started with informing the participants, getting their consent, filling out our demographic questionnaire, and giving them time to familiarize themselves with the VR setup and environment. After that followed the VR-based anatomy pre-test. The individual scores were written down by the study assistant. Then followed the VR learning session, this time, the participants were only given roughly 10 minutes. After that, the participants transitioned to the new consolidation phase where they had to correctly place 14 organs in the 3D Anatomy Atlas that were lined up on a table (see Figure 8 (left)). Then followed the post-test, again, mirroring the pre-test. Finally, the participants had to complete questionnaires about usability, motivation, and motion sickness and had the opportunity to give subjective feedback. The pre- and post-tests as well as the questionnaires were always done individually, while learning and consolidation were done collaboratively in the multi-user group.

6.3. *Results*

In this section, we report the results of our second study. To test for significant differences between the single-user and multi-user groups, we employed the independent samples t-tests when the data was normally distributed (most data) and the Mann-Whitney-U test otherwise (only the anatomy post-test).

6.3.1. *Demography*

This study consisted of (coincidentally again) $n = 33$ participants, divided into 11 for the single-user group and 22 for the multi-user group (11 pairs). We were careful to recruit participants who had not already participated in Study 1. The participants were predominantly in the age range of typical medical students with an average age of 30.4 years for single-users and 29 years for multi-users (standard deviation of 9.05 and 3.18, respectively). However, the participants encompassed a diverse array of university disciplines and had no specific medical training. We assume the anatomical knowledge level to be roughly similar to novice medical students, though. The single-user group comprised 8 male (73 %) and 3 female (28 %) participants, while the multi-user group comprised 20 male (91 %) and 2 female (9 %) participants (see Fig. 10 (left)). Previous VR experience among participants varied: In the single-user group, 18.2 % had no prior experience, 45.5 % had used VR before, and 36.4 % possessed significant experience. Comparatively, in the multi-user group, the distribution was 18.2 % with no experience, a larger 63.6 % having used VR previously, and 18.2 % with extensive experience, see Fig. 10 (right).

6.3.2. *Anatomy Knowledge, Learning Progress, and Time*

We measured mean pre-test scores of 16.0 for the single-user group and 16.909 for the multi-user group, with standard deviations of 3.256 and 4.286, respectively (see Fig. 11 (left)). The independent samples t-test yielded a p-value of 0.541. This indicates that there was no notable or even statistically significant difference in the baseline anatomy knowledge between the single and multi-user groups before the VR learning sessions.

After the VR learning session and the practical application phase, post-test scores improved to a mean of 25.545 for the single-user group and 26.818 for the multi-user group, with standard deviations of 2.697 and 1.918, respectively (see Fig. 11 (right)). Here, the post-test scores significantly deviated from a normal distribution, thus, we employed the Mann-Whitney U-test for statistical analysis. The resulting p-value was 0.087. Although this p-value is relatively small it does not quite reach the

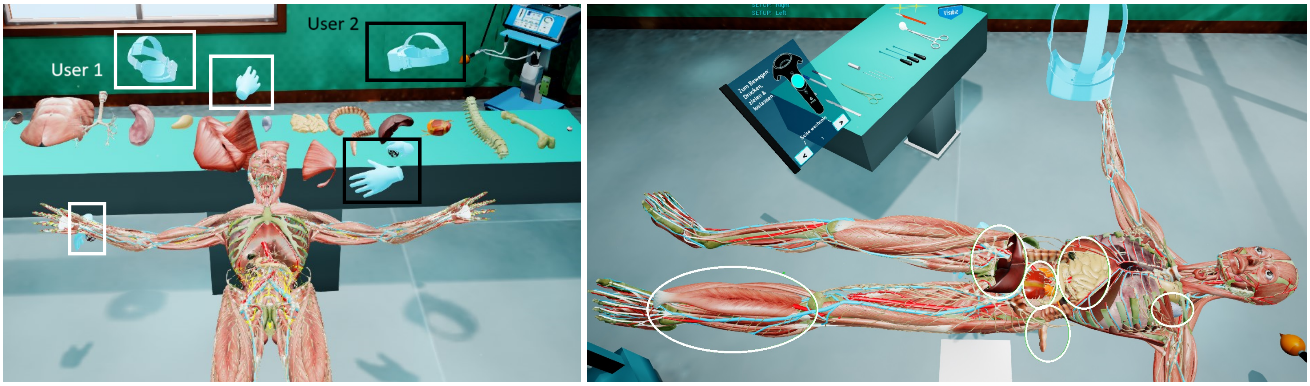

Fig. 8: Left: Two users (their avatars again highlighted with white and black boxes) within our added knowledge consolidation room/phase. The task is to correctly place the organs that lie on the table inside the human body. Right: A user conducting the newly designed pre-/post-test that consists of correctly identifying, naming, and relocating misplaced or superfluous organs (here highlighted in white circles.

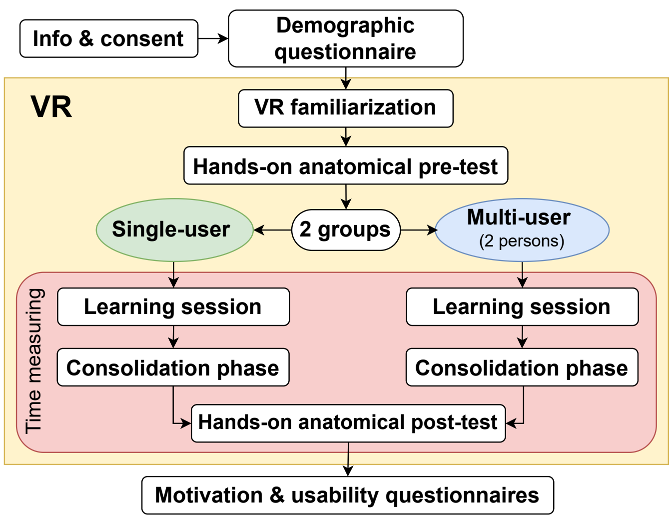

Fig. 9: Flow chart of our second study. Participants started with a hands-on anatomy pre-test in VR, followed by the learning and consolidation phases, both either individually or as a group. They then had to individually take the anatomy post-test, and complete additional questionnaires on paper.

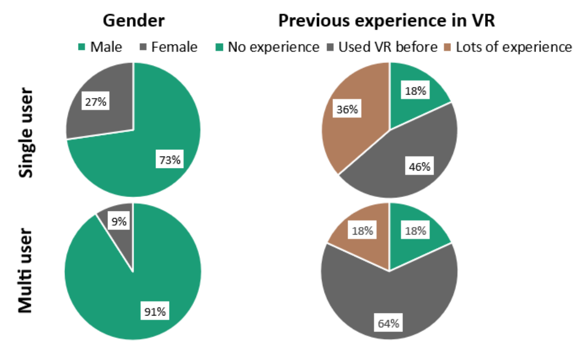

Fig. 10: Gender distribution (left column) and previous VR experience (right column) of the participants (in percent). The top row shows the single-user groups and the bottom row shows the multi-user groups. Most participants were men, especially in the multi-user group, and the latter had less VR experience, on average.

usual threshold of 0.005 for statistical significance. Nonetheless, there is at least a slight advantage for the multi-user group.

The learning progress (the difference between post-test and pre-test scores) was, on average, 9.545 for the single-user group and 9.909 for the multi-user group, with standard deviations of 4.083 and 3.878, respectively. The results are illustrated in Figure 12 (left). An independent samples t-test resulted in a p-value of 0.805. Thus, there seems to be no statistical difference between the groups regarding learning progress.

We found that the multi-user group, on average, spent more time learning, applying the knowledge, and completing the post-test (22.130 minutes) compared to the single-user group (18.011 minutes). The standard deviations were 3.665 and 3.556, respectively. The results are illustrated in Figure 12 (right). The independent samples t-test resulted in a p-value of 0.004, thus, the multi-user group did spend significantly more time in the VR session. We also measured the times the participants needed for the individual phases (see Fig. 13): In the learning session, multi-users, on average, spent more time (9.315 minutes) compared to single users (6.335 minutes). Similarly, for the consolidation phase, multi-users spent more time (7.003 minutes) than single-users (5.477 minutes). In contrast, for the post-test, single users and multi-users took roughly equally long (6.198 minutes, 5.812 minutes).

6.3.3. *Questionnaires on Motivation and Usability*

The analysis of the questionnaires on motivation, learning, and flow, see Fig. 14 (left), revealed the following mean scores: for single users, motivation was 4.571, learning 4.509, and flow 4.485; for multi-users, the scores were slightly lower with motivation at 4.273, learning 4.355, and flow 4.106. The standard deviations for motivation were 0.373 (single user) and 0.599 (multi user), for learning 0.404 (single user) and 0.474 (multi user), and for the flow 0.345 (single user) and 0.404 (multi user). Using the independent samples t-test, we found a significant difference in the flow scores between single and multi-users ($p = 0.012$), indicating that single users experienced a higher state of flow during the study. However, no significant differences were observed in motivation ($p = 0.142$) and learning ($p = 0.362$) scores between the groups.

The System Usability Scale (SUS) analysis yielded an average score of 83.409 for single users and 82.614 for multi-users, with standard deviations of 10.079 and 10.220, respectively, see

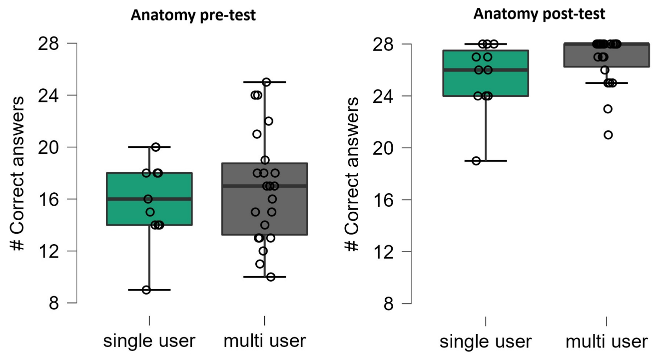

Fig. 11: Results of the anatomical knowledge pre-test conducted before the study (left) and the post-test conducted after the study (right). The pre-test scores are similar between groups, while in the post-test, the score of the multi-user group seems to be, on average, slightly higher than the one of the single-user group.

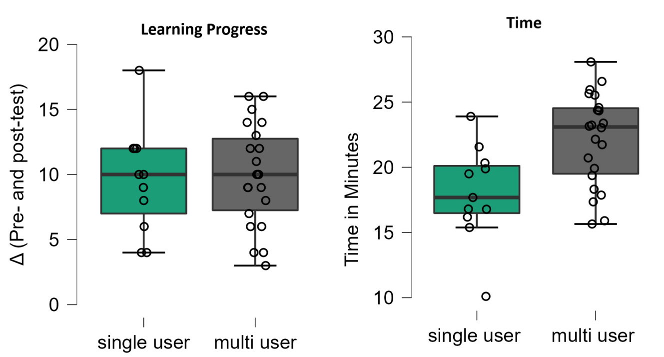

Fig. 12: Left: Learning progress (delta between pre- and post-test) for single- and multi-user groups. The seems to be no notable difference. Right: Time spent learning and consolidating the knowledge in VR. The multi-user group spent, on average, more time in VR.

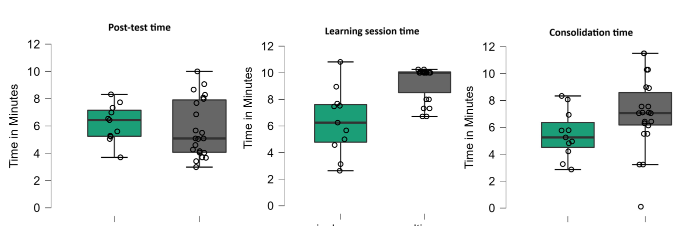

Fig. 13: Individual times for the times spent learning, consolidating the knowledge, and completing the post-test in VR. The multi-user group spent notably more time for learning and consolidation.

Fig. 14 (right). An independent samples t-test comparing the SUS scores between the two groups resulted in a p-value of 0.834. Thus, both groups rated the usability, similarly, as high.

7. Discussion

Although not the main focus of our work, we first wanted to verify whether our VR learning application and implementation are effective for anatomy learning in general, before investigating possible differences between single and multi-user groups. Looking at the results of our two studies, the post-test scores show a substantial improvement compared to the pre-test scores, for both groups, single-user and multi-user. Accordingly, on average, the participants had high learning progress. These positive results may come due to the VR learning environment allowing the participants to interact with the content in an immersive, engaging, and interactive way, which could have helped them better retain the information and recall it more easily during the post-test. The VR environment also allowed learners to visualize and explore anatomy in a three-dimensional way, which could have been helpful in understanding the subject matter and the spatial relations between anatomical structures. With these results, we can answer our first research question R_1: our Collaborative VR Anatomy Atlas is, generally, effective in enhancing the knowledge and understanding of anatomy. This result is consistent with previous research [7, 8], which found that learning in VR is beneficial.

Interestingly, the learning progress and post-test scores are, overall, not higher for the collaborative learning condition. In fact, in the first study, they tend to be slightly (but not statistically significant) lower than the ones for the single-user group. In the second study, the post-test scores for the multi-user group tended to be notably (but not significantly) higher, however, the learning progress was similar to the single-user group. Thus, we could not find VR learning to be more effective in collaboration than individually, which answers our research question R_2. This result is interesting as we would have expected advantages for the multi-user group since collaborative learning is generally considered beneficial [11, 12, 13].

A potential explanation for the higher (or at least similarly high) single-user learning outcomes in the first study could be that the single-user group had on average slightly less prior knowledge about anatomy (see the pre-test scores). This means that the single-user group had more learning potential. In the second study, the pre-knowledge was more similar between groups, and the post-test and learning progress scores of the multi-user group were at least more similar to or even slightly higher than the ones of the single-user group, which tends to support this hypothesis. This would be consistent with Dengel et al. [43], who found that prior knowledge influenced learning outcomes. Other possible reasons for the lack of significant benefits from collaboration across both studies could be that the participants may have felt more competition during the learning session and the VR post-test, or that our chosen avatar representation may not have provided a sufficient level of immersion, personalization, and embodiment, which possibly led to a low feeling of social presence. According to Dengel et

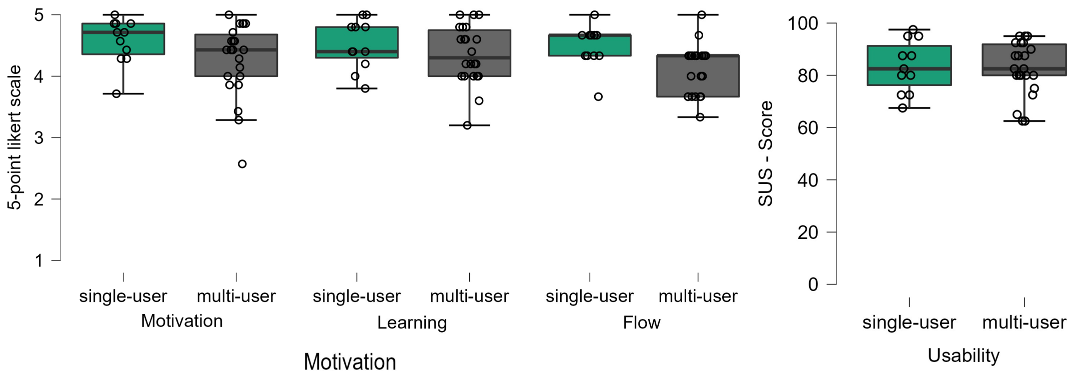

Fig. 14: Left: Results of the motivation questionnaire (factored into motivation, learning, and flow). The results are quite similar (high) for both groups, except for the flow factor, which is notably lower for multi-users. Right: The System Usability Scale scores, which are similar very good for both groups

al. [43] the level of presence does affect learning outcomes. Therefore, the potential benefits of collaboration may have not been fully exploited. The users' preferred learning setting could also have affected the learning experience and their resulting learning progress, since, at least in the first study, a substantially higher proportion of participants in the single-user condition reported preferring learning alone than in a group, while the ratio was more similar in the multi-user group. This may have influenced the results in favor of the single-user group for this study.

Furthermore, the participants who learned individually were potentially able to concentrate better on the task than the participants in the shared environment. The participants in the latter group were possibly more distracted by each other and the more complex multi-user environment, which provides additional social cues and stimuli (e.g., avatar and organ movements), and requires communication and coordination between users. According to collaborative cognitive load theory, this requires a considerable amount of cognitive resources and may have increased their cognitive load [44]. This may have been especially problematic considering that the virtual reality environment itself with its many visual objects might have already created a substantial amount of cognitive load, as suggested by Drey et al. [45]. High cognitive load can lead to lower learning outcomes according to Schild et al. [46], and, therefore, may have reduced the learning outcomes in our study. Previous work suggested that collaborative learning will be more effective when groups are heterogeneous, tasks require input and effort by all group members, and the advantage of dividing the task's intrinsic cognitive load between the group members outweighs the extra load by coordination and communication [44]. This might have not been the case in our study. Our results regarding the time spent learning in VR show that the multi-user group stayed significantly longer in the learning (and consolidation) sessions. On the one hand, this could be an indicator of a more engaging, positive user and learning experience, which multiple participants suggested after the study. On the other hand, the increased time may indicate slower learning progress and reinforce the assumption of higher cognitive load caused by a more complex, distracting environment for the multi-user group.

Moreover, it seems that the addition of the consolidation phase in the second study may have had a positive effect but did not lead to significantly better learning progress for the multi-user group, contrary to our assumption that this more practical learning activity might benefit more from collaboration. However, when looking at the post-test scores of Study 2, we see that the multi-user group fared slightly better than the single-user one (also approaching the usual significance threshold with $p = 0.087$) and that the scores were mostly close to the top of the scale. In fact, 13 out of 22 (59 %) multi-user participants scored the maximum points, while only 3 out of 11 (27 %) single users did. This could indicate that our post-test might have not been comprehensive enough, missing more and harder questions to accurately capture the knowledge learned. In turn, it might be possible that the collaborative group in Study 2 actually did have noticeable or even significantly better learning outcomes which we simply could not fully capture with our test. In this case, the added consolidation phase/changed task may have actually been effective in promoting collaborative learning. On the other hand, with the present results, we must also consider the possibility that the task of anatomy learning in VR may be one that does not automatically benefit from collaboration.

Regarding our research question R_3, we found that the results of the usability questionnaire are generally positive, especially for the second study. This shows that our Collaborative VR Anatomy Atlas provides a good user experience. The score for the multi-user group in the first study is noticeably lower than the one of the single-user group, though. This would reinforce our assumption that participants in the multi-user condition perceived the environment as more complex and demanding, potentially leading to a higher cognitive load (e.g., the need for coordination, and keeping track of the other person and what he is doing). Our chosen avatar representation (floating head+hands) may be also not ideal in this regard, too. Another possible reason could be that the participants in the group condition distracted each other, either simply by their presence and actions or by active off-task (group) behavior (i.e., testing the capabilities of the application, excessive socializing, etc.). The longer time spent in VR may be an indicator of this. In the end, the lower usability might be a central reason for the lower learning progress for the multi-user group, which would be consistent with the findings of Schild et al. [46].

In our second study, the usability scores were more similar between groups, as was the learning progress. Thus, usability might indeed be a relevant factor (lower usability might reduce learning progress), however, we found no evidence for it principally hindering collaborative learning specifically (even with equally high usability, the collaborative group did not have significantly better learning outcomes). Reasons for the usability being much higher in the second study and more similar between groups might be the time-limited, more focused (and enforced) learning session which could have helped to limit distracting behavior and undesired/unintended testing/exploration of the application and its functionality, especially in the multi-user condition. This would be in line with the guidelines by Drey et al. [45], who advised guiding the focus and attention of users on the task and relevant stimuli in collaborative learning sessions to reduce cognitive load and increase the user experience, presence and learning outcomes. Also, the changed pre- and post-tests might have increased the overall usability by being a more hands-on practical activity than reading text questions in VR and answering a virtual multiple-choice text. It seems that to achieve high usability, it is important to have a streamlined application, from a technical point of view, but also in terms of procedure/flow, that strongly encourages task focus, provides hands-on/enjoyable tasks, and avoids too much reading in VR.

We also got very positive average scores in our motivation questionnaire for both studies and both the single-user and multi-user groups, which indicates high levels of motivation and engagement when using our Collaborative VR Anatomy Atlas. The scores for all three factors (motivation, learning, flow) were mostly similar between the groups, although there is a tendency for the single-user scores to be slightly higher, especially regarding the flow factor. This answers our research question R_4. The learning and flow factors being higher, absolutely speaking, in the second study might be attributed, again, to the changed pre- and post-tests, the time-limited and more focused learning session, or the added consolidation phase. The overall motivation results are somewhat surprising, as, after analyzing the other results, we would have expected the multi-user group to fare worse in the first study and better in the second one to be in line with the usability results and our theories about the more complex, distracting shared environment and increased cognitive load. However, Du et al. [18] also found that single-player and multi-player learning groups had equally high motivation scores (even though the latter felt reportedly more stress). Hence, it seems that motivation might be more related to the use of a VR system rather than whether or not one is learning collaboratively. Moreover, for the first study, we found notable (but not significant) correlations between age and motivation (higher motivation with higher age), and between gender and motivation (higher motivation for women). The subjective feedback given by the participants during and after the learning session was generally very positive, too. They found the Collaborative VR Atlas to be an effective, useful, and enjoyable anatomy learning tool. Especially the collaborative scenario was often noted to provide an engaging, fun user and learning experience. Interestingly, this seems to not have translated to better scores in the questionnaires. Moreover, no participant experienced any signs of motion sickness.

These results - no benefits for collaborative learning in VR - are surprising and intriguing, as previous research has shown that VR systems are beneficial for learning [24], and that collaborative learning in non-VR settings is also beneficial [11, 15], e.g., higher gross anatomy learning outcomes [14]. Hence, the common assumption is that collaborative learning should also lead to better results than individual learning when conducted in VR. Wang et al. [31] recently showed that VR-based cooperative anatomy learning was superior to traditional learning, they did no direct comparison with individual VR-based learning, though. Our results indicate, however, that the benefits of traditional collaborative learning do not easily translate to VR. Apparently, some VR-specific factors, which are not fully explored yet (e.g., the specific task, usability, presence, avatars, etc.), have to be considered and taken into account in order to hopefully exploit the potential of collaborative learning in VR, too. For instance, we found that the usability and the task seem to be relevant, which is in line with previous studies [46]. However, increasing the former and changing the latter (more hands-on, closer to real procedure) did not automatically lead to significantly better collaborative performance. Naturally, one has to be cautious when generalizing based on these results, as our studies had only moderate sample sizes.

8. Limitations

One limitation of our studies is related to the adapted CMELAC questionnaire. Although we assume it to be still valid and reliable, we did not formally validate it. In addition, the number of participants was relatively small in both studies and usability ratings were mediocre in the first study (but good in the second one), so the study's power may be limited. Although we deliberately chose the head+hands avatar representation (easy-to-use, robust), it might have had a limiting effect on the perceived embodiment and (social) presence, which are especially relevant in the multi-user condition. Also, it might be possible that the new pre- and post-test variant in Study 2 might have had positive effects on usability and learning progress, but might have not been comprehensive enough to fully measure the latter (many participants of the multi-user condition with full scores).

9. Guidelines

In this section, we provide a set of guidelines and recommendations that we have formulated based on our findings to help guide the design of future VR learning tools.

- The benefits of traditional collaborative learning methods do not seem to transfer easily to VR, as application design and several VR-specific factors seem to have a significant impact on effectiveness.

- Usability, cognitive load and possibly (social) presence seem to be such important factors that need to be considered for effective collaborative VR learning. In general, more research is needed in these areas, e.g. whether more immersive, high-fidelity, full-body avatars would have positive effects.

- Minimize in-application distractions/non-task behavior in VR and guide participants' focus. For example, by hiding irrelevant aspects and information, highlighting the important ones, introducing time limits, etc.

- The learning task and assessment method should be hands-on; avoid multiple-choice questionnaires and text questions in VR.

- The test to measure learning progress should be comprehensive and difficult enough to fully capture progress and potential differences between groups.

- A hands-on consolidation phase such as in our Study 2 may be helpful, but more research is needed before definitive answers can be given.

10. Conclusions

To investigate the effectiveness of collaborative (anatomy) learning in virtual reality, we developed the Collaborative VR Anatomy Atlas, a multi-user VR application for anatomy education. We then conducted two user studies, each with a total of $n_{1,2} = 33$ participants, to compare individual learning with collaborative learning. Our application is built on Unreal Engine 4 and offers an immersive multi-user learning environment in which users can interactively and intuitively explore detailed human anatomical structures in 3D, including a model of the circulatory system.

The results demonstrate that our Collaborative VR Anatomy Atlas is effective for anatomy learning in both single and multi-user scenarios. Additionally, participants found the learning experience engaging and motivating, and reported moderate to high usability scores. However, no significant advantages or differences were found for the collaborative learning scenario in terms of learning effectiveness in compared to single-user learning. The results for motivation were generally similar, with a slight tendency for the collaborative scenario to score lower. Usability was slightly lower in one of the two studies, too, but was similar for both scenarios in the second study.

Our results show that the usual benefits of classical collaborative learning do not easily translate to VR environments. We suspect that the lack of advantages for the collaborative scenario is due to the more complex shared environment and higher cognitive load. Other possible reasons include the avatars used not being immersive enough, resulting in low social presence, or the fact that learning anatomy in VR may not necessarily benefit from collaboration. However, in our opinion, collaborative VR settings can be effective and efficient for learning complex spatial knowledge, similar to real-world learning situations. Nevertheless, our results indicate that further research is necessary to determine the most effective 3D interaction techniques and forms of collaboration in VR to achieve these goals.

11. Future Work

In the future, the first aspect we plan to focus on is to further improve the usability and user experience, especially for multi-user usage, by further streamlining the application and its flow and by developing improved interaction techniques and integrating more comfort features, such as different colors for each avatar/its hands. In order to do so, it would be important to investigate the level of usability in the individual phases of the study to better understand which part might need improvement. It also seems to be very important to actively encourage participants to focus on the desired task and to prevent non-task behavior, so further research should also investigate how to guide the focus, detect undesired behavior (e.g. using eye tracking), and how to prevent it (technically and through adapted procedures, e.g. time limits and achievements). A second important aspect is to adapt and improve the pre- and post-test to make it more comprehensive and to ensure that it is able to fully track learning over a wide range.

In general, it would be highly important that future studies are conducted with more participants and also include more (novice) medical students. This would help to enhance the generalizability of the findings. In addition, it would be important to normalize learning progress on the basis of pre-existing knowledge and to formally re-validate the adapted CMELAC questionnaire. We also propose to conduct further studies to examine the impact of presence, cognitive load, more realistic full-body avatars, and co-located vs remote collaboration. Future research could also focus on exploring larger group sizes and potential correlations between learning progress and gender mix within groups, on potential positive effects of adding a teacher/supervisor to each learning group, or on differences in long-term knowledge retention. Finally, it may be interesting to compare traditional VR setups with mixed reality setups.

Acknowledgment

We express our gratitude to Dr. Verena Uslar and Prof. Dr. med. Dirk Weyhe for their assistance with the anatomy-related questions and tasks. The software used in the study was built on top of the virtual anatomy atlas by Gloy et al. [25].

References

[1] Singal, A, Bansal, A, Chaudhary, P. Cadaverless anatomy: Darkness in the times of pandemic covid-19. 2020.

[2] Keyes, O, Peil, B, Williams, RM, Spiel, K. Reimagining (women's) health: Hci, gender and essentialised embodiment. ACM Transactions on Computer-Human Interaction (TOCHI) 2020;27(4):1–42.

[3] Guimarães, B, Dourado, L, Tsisar, S, Diniz, JM, Madeira, MD, Ferreira, MA. Rethinking anatomy: how to overcome challenges of medical education's evolution. Acta medica portuguesa 2017;30(2):134–140.

[4] Smith, J, Jones, S. Anatomy education: classical techniques and modern advancements. Medical Education Journal 2020;10(2):50–60.

[5] Estai, M, Bunt, S. Best teaching practices in anatomy education: A critical review. Annals of Anatomy-Anatomischer Anzeiger 2016;208:151–157.

[6] Bonali, M, Bernardeschi, D, Patel, N, Ghirelli, M, Fernandez, IJ, Fermi, M, et al. Surgical instruments and preparation of the specimen. Comparative Atlas of Endoscopic Ear Surgery: Training Techniques Based on an Ovine Model 2021;:9–27.

[7] Markant, DB, Ruggeri, A, Gureckis, TM, Xu, F. Enhanced memory as a common effect of active learning. Mind, Brain, and Education 2016;10(3):142–152.

[8] Roh, TH, Oh, JW, Jang, CK, Choi, S, Kim, EH, Hong, CK, et al. Virtual dissection of the real brain: integration of photographic 3d models into virtual reality and its effect on neurosurgical resident education. Neurosurgical Focus 2021;51(2):E16.

[9] Slater, M. Immersion and the illusion of presence in virtual reality. British journal of psychology 2018;109(3):431–433.

[10] Barfield, W, Zeltzer, D, Sheridan, T, Slater, M. Presence and performance within virtual environments. Virtual environments and advanced interface design 1995;:473–513.

[11] Casey, A, Goodyear, VA. Can cooperative learning achieve the four learning outcomes of physical education? a review of literature. Quest 2015;67(1):56–72.

[12] Johnson, DW, Johnson, RT, Stanne, MB. Cooperative learning methods: A meta-analysis. Minneapolis; 2000,.

[13] Kyndt, E, Raes, E, Lismont, B, Timmers, F, Cascallar, E, Dochy, F. A meta-analysis of the effects of face-to-face cooperative learning. do recent studies falsify or verify earlier findings? Educational research review 2013;10:133–149.

[14] VanLeuven, AJ, Szymik, BG, Ramsey, LM, Hesse, DW. A multi-year evaluation of medical student performance on and perceptions of collaborative gross anatomy laboratory examinations. Anatomical Sciences Education 2023;16(3):557–566.

[15] Rutherford, SM. Collaborative learning: Theory, strategies and educational benefits. Nova; 2014.

[16] Nokes-Malach, TJ, Richey, JE, Gadgil, S. When is it better to learn together? insights from research on collaborative learning. Educational Psychology Review 2015;27:645–656.

[17] Laal, M, Ghodsi, SM. Benefits of collaborative learning. Procedia-social and behavioral sciences 2012;31:486–490.

[18] Du, YC, Fan, SC, Yang, LC. The impact of multi-person virtual reality competitive learning on anatomy education: a randomized controlled study. BMC Medical Education 2020;20:343.

[19] Al Maree, H, Fischer, R, Weller, R, Uslar, V, Weyhe, D, Zachmann, G. Collaborative vr anatomy atlas investigating multi-user anatomy learning. In: International Conference on Virtual Reality and Mixed Reality. Springer; 2023, p. 122–135.

[20] Preim, B, Saalfeld, P, Hansen, C. Virtual and augmented reality for educational anatomy. In: Digital Anatomy: Applications of Virtual, Mixed and Augmented Reality. Springer; 2021, p. 299–324.

[21] Falah, J, Khan, S, Alfalah, T, Alfalah, SFM, Chan, W, Harrison, DK, et al. Virtual reality medical training system for anatomy education. In: 2014 Science and Information Conference. 2014, p. 752–758.

[22] Fairén González, M, Farrés, M, Moyes Ardiaca, J, Insa, E. Virtual reality to teach anatomy. In: Eurographics 2017: education papers. European Association for Computer Graphics (Eurographics); 2017, p. 51–58.

[23] Codd, AM, Choudhury, B. Virtual reality anatomy: Is it comparable with traditional methods in the teaching of human forearm musculoskeletal anatomy? Anatomical sciences education 2011;4(3):119–125.

[24] Kurul, R, Ögün, MN, Neriman Narin, A, Avci, Ş, Yazgan, B. An alternative method for anatomy training: Immersive virtual reality. Anatomical Sciences Education 2020;13(5):648–656.

[25] Gloy, K, Weyhe, P, Nerenz, E, Kaluschke, M, Uslar, V, Zachmann, G, et al. Immersive anatomy atlas: Learning factual medical knowledge in a virtual reality environment. Anatomical Sciences Education 2022;15(2):360–368.

[26] Reinschluessel, AV, Muender, T, Salzmann, D, Döring, T, Malaka, R, Weyhe, D. Virtual reality for surgical planning–evaluation based on two liver tumor resections. Frontiers in Surgery 2022;9.

[27] Kaluschke, M, Yin, MS, Haddawy, P, Srimaneekarn, N, Saikaew, P, Zachmann, G. A shared haptic virtual environment for dental surgical skill training. In: 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE; 2021, p. 347–352.

[28] Fischer, R, Chang, KC, Weller, R, Zachmann, G. Volumetric medical data visualization for collaborative vr environments. In: Virtual Reality and Augmented Reality: 17th EuroVR International Conference, EuroVR 2020, Valencia, Spain, November 25–27, 2020, Proceedings 17. Springer; 2020, p. 178–191.

[29] Zorzal, ER, Sousa, M, Mendes, D, dos Anjos, RK, Medeiros, D, Paulo, SF, et al. Anatomy studio: a tool for virtual dissection through augmented 3d reconstruction. Computers & Graphics 2019;85:74–84.

[30] Jorge, J, Belchior, P, Gomes, A, Sousa, M, Pereira, J, Uh, JF. Anatomy studio ii a cross-reality application for teaching anatomy. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2022, p. 441–444.

[31] Wang, CY, Yin, T, Ma, KH, Shyu, JF, Cheng, CP, Wang, YC, et al. Enhancing anatomy education through cooperative learning: harnessing virtual reality for effective gross anatomy learning. Journal of Microbiology & Biology Education 2023;24(3):e00100–23.

[32] Boedecker, C, Huettl, F, Saalfeld, P, Paschold, M, Kneist, W, Baumgart, J, et al. Using virtual 3d-models in surgical planning: workflow of an immersive virtual reality application in liver surgery. Langenbeck's archives of surgery 2021;406:911–915.

[33] Schott, D, Saalfeld, P, Schmidt, G, Joeres, F, Boedecker, C, Huettl, F, et al. A vr/ar environment for multi-user liver anatomy education. In: 2021 IEEE Virtual Reality and 3D User Interfaces (VR). 2021, p. 296–305.

[34] De Back, TT, Tinga, AM, Louwerse, MM. Learning in immersed collaborative virtual environments: design and implementation. Interactive Learning Environments 2021;:1–19.

[35] de Back, TT, Tinga, AM, Nguyen, P, Louwerse, MM. Benefits of immersive collaborative learning in cave-based virtual reality. International Journal of Educational Technology in Higher Education 2020;17:1–18.

[36] Sedlák, M, Šašinka, v, Stachoň, Z, Chmelík, J, Doležal, M. Collaborative and individual learning of geography in immersive virtual reality: An effectiveness study. PLoS ONE 2022;17:Article e0276267. doi:10.1371/journal.pone.0276267.

[37] Lee, JE, Han, IH, Kim, SH, Kim, KS. Effectiveness of virtual reality-based learning for anatomy education: A systematic review and meta-analysis. Anatomical Sciences Education 2021;14(5):587–600.

[38] van der Meer, N, van der Werf, V, Brinkman, WP, Specht, M. Virtual reality and collaborative learning: a systematic literature review. Frontiers in Virtual Reality 2023;4. doi:10.3389/frvir.2023.1159905.

[39] Epic Games Inc., . Developer documentation of the onlinesubsystemnull module for multi-player testing. 2024. URL: https://docs.unrealengine.com/4.27/en-US/API/Plugins/OnlineSubsystemNull/FOnlineSubsystemNull/.

[40] Perfecto, C, Elbamby, MS, Del Ser, J, Bennis, M. Taming the latency in multi-user vr 360°: A qoe-aware deep learning-aided multicast framework. IEEE Transactions on Communications 2020;68(4):2491–2508.

[41] Brooke, J. Sus: a retrospective. Journal of usability studies 2013;8(2):29–40.

[42] Manzano-León, A, Camacho-Lazarraga, P, Guerrero-Puerta, MA, Guerrero-Puerta, L, Alias, A, Aguilar-Parra, JM, et al. Development and validation of a questionnaire on motivation for cooperative playful learning strategies. International Journal of Environmental Research and Public Health 2021;18(3):960.

[43] Dengel, A, Maegdefrau, J. Immersive learning predicted: Presence, prior knowledge, and school performance influence learning outcomes in immersive educational virtual environments. 2020,.

[44] Janssen, J, Kirschner, P. Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda. Educational Technology Research and Development 2020;68:1–23.

[45] Drey, T, Albus, P, der Kinderen, S, Milo, M, Segschneider, T, Chanzab, L, et al. Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22; Association for Computing Machinery; 2022,.

[46] Schild, J, Lerner, D, Misztal, S, Luiz, T. Epicsave – enhancing vocational training for paramedics with multi-user virtual reality. 2018,.

Appendix A. Additional Images

Fig. A.15: Animated 3D model of the circulatory system including both the arteries (red) and the veins (blue).

Fig. A.16: Interactive medical instruments (e.g., laser pointer, markers, scalpel, etc.) and the virtual tablet (e.g., to dynamically hide structures such as bones, change settings, and navigation between levels).

Fig. A.17: The consolidating phase/room in which the participants have to correctly place the organs lying on the table into the human body to practice and consolidate the new knowledge.

Appendix B. Questionnaires

Appendix B.1. Demographics questionnaire

Gender:

- Male
- Female
- Others

Age: ______
Profession: ______________________

Which learning type are you?

- Visual type
- Auditory type
- Haptic type
- Intellectual type

Do you usually learn more efficiently?

- Alone
- In a group
- Both

Previous experience in VR:

- none ○ ○ ○ a lot

Have you ever tried a VR learning programme?

- Yes
- No

Appendix B.2. Adapted Motivation Questionnaire (CMELAC) used in both studies (7-point Likert scale from "totally disagree" to "totally agree"):

1. In general, I have enjoyed this playful activity in VR Anatomy Atlas.
2. I would repeat these types of activities.
3. I have felt motivated.
4. I improved my knowledge of Anatomy.
5. My interest in the subject of Anatomy has increased.
6. My interest in learning in the VR environment has increased.
7. This activity format has been appropriate to check my knowledge of the subject (Anatomy).
8. Helped me identify my weaknesses in the subject.
9. It helped me understand the content of the subject.
10. With these types of activities, I learn more than in traditional classes.
11. I enjoyed the interaction (e.g., grasping organs in hand, using medical instruments) in the VR environment
12. The interaction elements have motivated me to carry out the activity.
13. While playing in a VR environment, I was not aware of what was happening around me.
14. I felt capable of carrying out the proposed activities.
15. I found the activities comforting and valuable to me.

Appendix B.3. System Usability Scale (SUS) used in both studies (5-point Likert scale from "strongly disagree" to "strongly agree"):

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.

5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Appendix B.4. Questions of the Pre-/Post-test for Study 1:

Where is the pancreas situated?

- In front of the stomach
- Behind the stomach
- To the left of the spleen
- Below the large intestine
- Behind the kidney

Where are the adrenal glands situated?

- Above the kidneys
- In front of the kidneys
- Behind the kidneys
- In the lesser pelvis
- Above the diaphragm

To what is the Medulla oblongata (extended spinal cord) connected?

- Spinal chord nerves
- Back muscles
- Rip muscles
- Abdominal muscles
- Neck muscles

How many main vessels are responsible for supplying blood to the hand?

- 1
- 2
- 3
- 4
- 5

Where in the abdomen is the appendix situated?

- Right part
- Lower left part
- Lower part
- Lower right part
- Upper part

Where is the spleen situated?

- To the right of the stomach
- To the left of the stomach
- In front of the pancreas
- Behind the left kidney
- Above the left kidney

Where are the lymph nodes of the neck situated?

- In front of the neck muscles
- Superficially and deep within the neck
- Behind the neck muscles
- Behind the windpipe
- In the spinal canal

How many muscles make up the frontal thigh muscles?

- 2
- 3
- 4
- 5
- 6

Appendix B.5. List of misplaced organs and grading criteria used in Pre-test & Post-test of Study 2:

Criteria:	Incorrectly placed, Organ name, Correct location, This instance superfluous
Organs:	Heart (Cor), Liver (Jecur), Stomach (Ventriculus), Spleen (Splen), Larynx Trachea Bronchus (windpipe), Small Intestine (Intestinum Tenue), Large Intestine (Intestinum Crassum), Muscle thigh (Left Sartorius - Left Vastus Medialis - Left Rectus Femoris), Right Urinary System (Systema Urinarium Dexter)

Appendix C. Raw Data

Appendix C.1. Subjective feedback given at the end of User Study 1:

- The introduction helped me in VR, without the introduction I answered a question in the quiz twice.
- For small objects, it would be useful to have a button to reset them to their original position (regardless of the area already selected).
- Since the system is designed for medical students, it's great.
- Grapping organs in hand and learning about muscles in this way really made me more interested in learning anatomy through this VR setup. Improvement:maybe the connected muscles information would be nice to have.
- Thanks for your great work!
- Make the body rotatable. Simplify tablet interactions. While moving something, highlight the place where it comes from.
- I really enjoyed it!
- Possibly the design of the questionnaire at the end, for example to match the colours of the operating room or the other surroundings.
- It could help if you couldn't reach through the upper layers.
- Very well illustrated.
- Touching and moving the organs helped to understand their position in the body better than 2D images. It was helpful to have a conversation with someone else about things like what kind of organ it is. I didn't feel sick, just dizzy, but it was a bit disorientating that the other person wasn't in the VR.
- Activity is good.
- If more than one person is playing at the same time and one person changes the settings, the other person's settings may change.
- Buggy tablet in multiplayer.
- Optimisation of the simulation with two players at the same time.
- Dizziness after use and dry eyes.
- It was fun. After the activity I was slightly dizzy and the perception (visually) is a bit strange.
- I would have liked to be able to hide the muscles.
- Better learning results when learning on your own or with a teacher.
- When placing an organ (bone or muscle) back, it should always be shown where it was originally supposed to go, even if it is far away.
- Gravity cool!

Appendix C.2. Subjective feedback given at the end of User Study 2:

- [heart symbol]
- Good resolution and 3D organs.
- The VR Activity to understand Anatomy was extremely fun. If I can study like this for everything...
- Thank you, it was my first time for this experience (Symbol happy face).
- I think there should be Wiki-like information about the organs if we triple touch them.
- The control of the positions of organs are not determined by their own abilities. The examiner was necessary. Accordingly, you are very dependent on him. If you want to learn 'on your own', it could be difficult to determine the positions without prior knowledge.
- I didn't realise my own mistakes. I probably wouldn't have improved without the advice from the examiner.
- Maybe start without nerves and vessels to reduce clutter.
- Multiuser was laggy and sometimes imprecise. Headset causes headaches.
- VR headset like Vision pro would be better. VR headset is uncomfortable, I don't like wearing it.
- Very small body parts were sometimes difficult to grasp. Otherwise, the system was easy to learn and to operate/use. The interaction with the organs was fun and helped me to better localise the latter in the body.
- I think the principle of learning anatomy is good. A magnifying glass would be great for finding hidden organs.
- For a long time I felt a bit dizzy. Overall, however, I felt well/comfortable.
- VR makes me headachy, apart from that I love it.
- It was fun and very informative. There were a few things I didn't know and where I was able to expand my knowledge.