Adaptive Training of the Mental Rotation Ability in an Immersive Virtual Environment

Sunita Ariali (✉), Bernd Zinn
University of Stuttgart, Stuttgart, Germany
sunita.ariali@ife.uni-stuttgart.de

Abstract—Virtual reality (VR) opens new possibilities for the investigation and training of the mental rotation ability (MRA), which is an important factor in the development of technical skills in several fields. Adaptive designs of MRA training environments realised by virtual technology, which are investigated in this study, could offer promising potentials. For the evaluation of the effectiveness, the adaptive training environment is compared with the adequate randomised environment by assessing the mental rotation ability in both conditions before and after training. As a dependent variable, the performance and its improvement in the virtual mental rotation test (VMRT), as well as the cognitive load, are measured. In addition, possible gender differences and their influence on the training outcomes are determined. The study described here represents an innovative support option for MRA and provides an expandable empirical basis for VR-based adaptive trainings.

1 Introduction

Virtual realities (VR) are becoming increasingly more important for learning and educational purposes [1]. The reasons for this are the technological developments of recent years on the one hand, and VR's property of being applicable in many different contexts on the other hand. Examples of VR use in educational contexts are virtual laboratories that support the acquisition of certain skills and practical experience [2] or the possibility of training in environments that would be too dangerous in reality [3], [4]. A special characteristic of VR is the egocentric representation of the environment and the use of a head-mounted display (HMD) which gives the user the feeling of being in "real" reality. Under certain circumstances this can increase the user experience, perceived competence, benefits, and motivation of learners [5]. In addition, immersive virtual technologies enable the use of natural user interfaces (e.g., Leap Motion, data gloves, etc.) to interact with virtual objects and imitate a real haptic touch. Virtual environments also offer a wide range of design options for training and learning scenarios. For example, there are options for the adaptive design of the training environment regarding the difficulty of the content being customised during the training process to reflect the skills, preferences, motivation, or other characteristics

of the student. This can reduce the cognitive load while improving the overall performance [6].

In light of the advantages described above, VR offers the possibility of creating training sessions that require realistic interaction with the environment while being equipped with advanced features at the same time. The training of the mental rotation ability (MRA) is one of these possibilities. MRA, as an important part of the spatial imagination and thus of human intelligence, is a prerequisite in many scientific and technical professions [7]. Previous research has shown a marked gender difference to the detriment of women [8], [9]. Nonetheless, MRA can be supported by means of appropriate training programmes [10]. The possibility to train MRA in VR with the help of adaptively designed materials offers an interesting research field in this regard and holds practical potentials.

This study investigates the effectiveness of VR-based adaptive training, which was developed to support MRA. During training, the test subjects had to detect the similarity of three-dimensional figures and rotate them into the correct positions. In order to ensure the adaptivity of the training environment, the next task became more difficult if the previous task was solved correctly (if the upper difficulty limit had not already been reached) and easier if it was not solved correctly (if the lower difficulty limit had not already been reached). Adaptive training was compared with randomised training, in which the training tasks were presented in no particular order.

In order to test the effectiveness of training, a previously validated virtual test instrument for measuring MRA [11], the so-called virtual mental rotation test (VMRT), was used before and after training It was assumed that the participants will achieve better results in VMRT after the virtual training than before the training if the training has a positive influence on MRA This improvement should be stronger in the adaptive condition than in the randomised condition. Additional tests examined the cognitive load in different conditions as well as gender-specific differences. Following the introduction, this paper will firstly provide an overview of the theoretical background and related work. Subsequently, the research questions of the current study are formulated. Next, the methodological approach of the study is explained in the third section, followed by a report on the results in the fourth section. In the fifth section, the central findings are discussed before finally summarising the study.

1.1 VR technology

The rapid development and dissemination of new technologies in recent decades is linked to the optimisation of educational processes. With the integration of digital technologies, learning and work processes can be improved by making them more flexible and individualised. One of these technologies is VR, which is no longer a novelty but has experienced a strong developmental thrust and spread in recent years. VR is a computer-generated three-dimensional (3D) representation of real or fictitious environments in which the user can enter and execute nearly realistic actions. Modern technologies offer the possibility to "dive" into these environments and get the feeling of actually being "there". This feeling is described as immersion, which Freina and Ott [12] define as the feeling of being physically present in a non-physical world.

Depending on the degree of immersion, currently existing VR systems are divided into three categories: non-immersive (desktop), semi-immersive, and fully immersive systems [13]. In contrast to the desktop-based representation (non-immersive technology), the special feature of immersive virtual environments is that the user can view the surroundings from an egocentric, i.e. "I" perspective, which corresponds to a realistic perception of the environment and its objects [14]. This is usually enabled by using an HMD, also known as VR goggles. While fully immersive VR technology enables a wide range of interactions with virtual objects, the possibilities for interaction in semi-immersive environments are limited as they are mostly realised by means of low-budget devices as well as smartphones and contain only few natural user interfaces. Technological examples for the realisation of semi-immersive environments are Google Cardboard or the VR goggles Samsung Gear. The multiple possibilities for interaction with fully immersive technology (i.e., HTC VIVE) is provided by an HMD connected to the PC, space-scanning infrared cameras, and special controllers. Via the controllers, the users can interact with the virtual 3D objects, take actions, and move within the virtual space. In order to enhance the feeling of immersion, VR technology can be extended with additional natural user interfaces. As an example, the Leap Motion technology, which is attached to the front of the HMD, can track the user's hands and transfer them as models into a VR environment. In this case, controllers are no longer needed, as users can interact with the virtual objects by means of their own hands. Although fully immersive virtual environments offer numerous advantages for the user compared to non-immersive or semi-immersive environments, they are not always the means of choice for economic reasons. Fully immersive VR technology, such as HTC VIVE, is quite expensive, requires a powerful PC, and the use of a room-scanning infrared camera is cumbersome. In addition, programming realistic 3D models and integrating interaction options is complex and requires programming skills, even though current game engines, such as Unity3D or Unreal Engine, use a visual scripting system and an intuitive editor to build simple games without writing code. Many finished scripts are also available as open source.

Despite the high costs and efforts involved, the importance of virtual technology in the education sector is so high that it is even described as the number one learning tool of the 21st century [15]. This is due to its unique characteristics, providing several advantages for education and research. As an example, VR allows for the simulation of realistic but fictitious environments that would be impossible or too dangerous under real conditions [16]. Furthermore, VR enables the modification of parameters that often cannot be changed in a real system [2]. Examples of modifiable parameters are gravity, colour, or size, which can be set by default or varied according to individual preferences. An additional advantage of using computer-generated environments is the ability to automatically log and analyse study-relevant data, which is an efficient alternative to other data collection methods [17]. VR also offers the possibility to take the collected data into account and adapt the learning environment to the needs of the user.

The above-mentioned advantages of virtual environments, in combination with years of research results, provide an opportunity for using VR to train certain skills or as a supplement to real or computer-based training environments. Such a form of

training has partly already been implemented [1]. However, it requires more detailed research as well as the identification of weaknesses and strengths in order to enable extensive use. Applications of virtual environments for training spatial skills have also often been addressed in research. With respect to spatial orientation, there is a number of studies that show the benefits of virtual environments for the training of spatial skills [18], [19]. The researchers assume that, in order to orientate themselves in a virtual environment, people rely on the same cognitive process as they do in a real environment [20]. While graphically complex and expensive virtual environments, in which walking around and exploring the surroundings is recommended, are suitable for spatial orientation, a further dimension of spatial abilities, the mental rotation ability, can be assessed with simple stimuli. Therefore, especially MRA is frequently examined in connection with virtual environments and is thus described in the next chapter.

1.2 Mental rotation ability

The term MRA refers to the ability to rotate 2D or 3D figures and objects mentally [21]. It is allocated to the realm of spatial abilities which are an aspect of human intelligence [22]. Spatial ability means being capable of storing, retrieving and transforming visual-spatial information [23]. There are various theories and models of spatial abilities. Thurstone [24], for example, postulates a three-factor hypothesis according to which visualisation, spatial relations, and spatial orientation are the three most important subfactors of spatial ability. Another important definition of spatial ability comes from Linn and Petersen [21], who also identify three subfactors of the construct, including spatial perception, mental rotation, and spatial visualisation. Carroll [25] divides spatial abilities into five main factors, comprising spatial visualisation, spatial relationships, closing speed, the flexibility of closing, and perceptual speed. Despite these discrepancies, mental rotation is a component of all these models and theories. It is the ability to mentally rotate the representation of a stimulus in order to imagine what an object looks like from a different perspective [26].

A frequently used diagnostic possibility for measuring MRA is the mental rotation test (MRT), according to Shepard and Metzler [26], in the version by Peters et al. [27]. The 24 MRT items consist of 3D cube constructions. For each test item, a given figure is compared with four other figures on the right-hand side. Two figures on the right are rotations of the target figure, and the other two figures are not identical to the target figure. The two rotated versions should be recognised by the test persons. Then, and only then, if both versions are correctly recognised, one point per item is awarded so that the maximum score in the MRT is 24.

The research indicates the importance of MRA for technical professions [28]. Sorby et al. [29], Sorby and Baartmans [30] and Veurink and Sorby [31], for example, show that improving spatial skills, especially MRA, can lead to a better performance in mathematics, computer science, science, and technology. Other authors also find significantly high positive correlations between mathematical skills and spatial abilities [32], [33].

An important characteristic feature of MRT is gender differences [9]. With the paper-and-pencil MRT of Shepard and Metzler [26], the disadvantage of women compared to men in the realm of mental rotation was detected for the first time and has been proven several times ever since. Modern computer-based MRT also confirms these results [34]. Gender differences in spatial abilities are well documented but have not yet been adequately explained. Various explanations are given, which are based on biological [35], [36], environmental [37], or psychobiological aspects [7]. One of these explanations refers to the way the tasks are presented [38]. The paper-and-pencil or screen-based representations of the MRT are exclusively allocentric. There is evidence that women have difficulties with such representations. Consequently, if the mental rotation figures were represented from the egocentric perspective, they would perform as well as men [39]. With VR technology, new presentation possibilities for MRT are emerging that did not exist before or that were technologically complex and expensive. With the help of immersive technology, Shepard's and Metzler's [26] stimuli (cube figures) could also be viewed from an egocentric perspective. This opens new possibilities to fundamentally investigate spatial imagination and gender differences in MRT tasks.

Due to its importance in many areas, the support of mental rotation constitutes an important area of research. Meanwhile, there is ample evidence that MRA can be improved by training [10], [40]–[42] and that the effects of training can be transferred to other tasks [43]. According to some authors, cognitive processes underlying mental rotation are linked to actual physical rotation and can activate corresponding motor processes. Wohlschläger and Wohlschläger [44] propose that mental and manual rotations are based on a common process. The authors point to a number of studies that found a correlation between the reaction times of manual and mental rotations. Based on this common process hypothesis, it can be assumed that manual training is able to improve MRA. This assumption is supported by a study by Wiedenbauer et al. [41] in which the authors develop a training task requiring the participants to align the orientations of two Shepard and Metzler stimuli using a joystick. The manual training resulted in a better mental rotation performance with familiar objects. Meanwhile, there is further evidence for the effectiveness of manual training for MRA [40].

The realisation and investigation of manual training environments of MRA are challenging in real conditions because of the difficult standard operability and control of environmental confounding variables. The immersive virtual realities, in combination with data gloves or Leap Motion technology, make it possible to realise training environments that enable the elimination or control of the disturbing variables and allow the standardisation of relevant parameters. Additionally, virtual environments offer the possibility to adapt the learning content to the skills and needs of the users by means of automated evaluation of the data. This kind of designing and learning environments is called adaptive and is described in detail in the next section.

1.3 Adaptive training environments

Adaptive learning is a kind of learning process in which the learning content is changed or adapted based on the reactions of the learner [45]. Accordingly, adaptive

training can be described as training that matches individual differences with training instructions [46]. In an adaptive training scenario, the order, pace, or difficulty of the content can be tailored individually during the training process – for example, to the skills, preferences, or motivation of the student – in order to increase training efficiency [47]. The concept of adaptive training is very broad, and there are already many realisations accompanied by a long research tradition. Research indicates that adaptive training can be effective in general when compared to non-adaptive training [48], [49]. The effectiveness of adaptively designed training environments can be supported by the cognitive load theory (CLT, [50]). This theory is based on the human memory architecture model of Baddeley [51] and assumes that the working memory capacity is limited. The CLT distinguishes three types of working memory loads. Sweller [52] describes the load caused by the type of learning material as intrinsic load. This load increases with the number of elements that are simultaneously present in the working memory, whereby the extent of the intrinsic load also depends on the level of expertise of the learner. The second load specified in the CLT is the extrinsic load. This can be directly influenced by the instructional design of the learning material. The cognitive effort involved in automating schemata also results in the learning-related load (germane load). The goal of CLT-based training is to increase the germane load and decrease the extraneous load. According to CLT, the use of adaptive training should relieve working memory and promote germane load, as it is tailored to the individual needs and abilities of the learner [6].

Personalisation of virtual environments entails a number of advantages which are known in the gaming industry [53]. Adaptability can be used to prevent learners from being overloaded or too distracted by irrelevant content, which could increase the flow experience. Additionally, the learning effect could be increased by adapting the complexity of the content presented in virtual reality according to the learner's abilities.

The idea of adaptively designing learning materials presented in virtual form is not new [54], but research in this area is still in its infancy [55]. The aim of this study is to examine a simple form of personalised training and compare it with non-personalised training. The variables of gender and cognitive load are going to be considered and included in the regression analysis.

2   Current Study

Based on the state of the research described above, an adaptive design of the virtual training environment to promote mental rotation has a high practical relevance and raises new research questions. First, it has to be questioned whether virtual training can improve MRA. In order to investigate this, this study uses a pretest-posttest design examining whether an increase in performance is achieved. It is expected that a significant improvement in MRA will occur after a virtual training session. Therefore, the first hypothesis is as follows:

H1: A higher MRA can be measured after training than before training.

Next, it has to be checked whether the adaptive design of the training environment has an advantage over the randomised environment, which serves as a control condition in this study. It is assumed that training in the adaptive condition leads to a higher increase in MRA than in the randomised condition. This assumption is based on the circumstance that in an adaptive environment, the difficulty of the given task is determined by the person's solution behaviour, thus avoiding over- or underchallenging the person. In addition, the person to be trained has the possibility to solve the mental rotation tasks within his or her own scope in order to achieve better training effects. Therefore, in the second hypothesis, it is assumed that significantly higher increases in MRA will be measured after adaptive training than after randomised training. Therefore, the following hypothesis is established:

H2: The increase in ability is higher in the adaptive condition than in the randomised condition.

Given the adaptability of the task difficulty to the abilities of the user, it could be hypothesised that the overall cognitive load in the adaptive condition is lower than in the randomised condition. Under certain circumstances, this should also be measurable with subjective measurement methods after the virtual phases. For this purpose, the cognitive load is assessed in addition to MRA and checked for group differences. Hence an additional question follows:

Do different groups differ in cognitive load?

Considering that gender is an important variable in the context of mental rotation, it is investigated whether gender differences can be identified by virtual MRT and by cognitive load (measured by NASA-TLX), and how they develop under different training conditions. From this, the supplementary question of the study is derived:

Are there gender differences in relation to performance in virtual MRT and/or cognitive load?

3 Method

3.1 Participants

102 persons participated in the study. 51 of them (50%) were female, and the other 51 (50%) were male. The mean age of the participants in the study was 27.74 years with a standard deviation of 5.66 years (minimum 18 years, maximum 47 years). 47 of 102 participants already had a university degree and 30 participants had a high school diploma. The remaining participants had other degrees. When asked how often they used 3D computer games, 67 of 102 test persons stated "never", 18 "very rarely", 12 "occasionally" and 4 "almost daily." One person did not provide any information in this regard.

3.2 Design

The study consists of a pretest-posttest experimental-control group design. The adaptive training serves as an experimental condition and the randomised training as a control condition. Before and after the training session, the MRA of the participants is recorded using a virtual MRT which was developed and validated in the previous study [11]. Finally, the cognitive load of the persons is recorded using the NASA TLX questionnaire. Figure 1 shows a graphic representation of the course of the investigation.
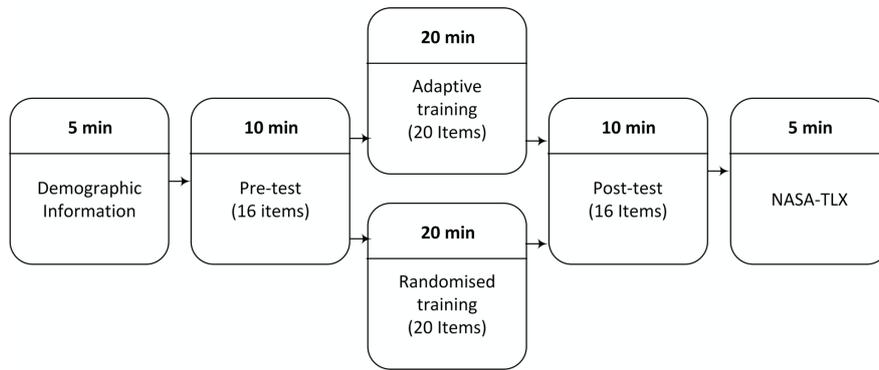

Fig. 1. Course of the study session

3.3 Materials

The materials used in the study were VR-based mental rotation test and training environments, the technology required to run the VR environments and the NASA TLX questionnaire as well as the questionnaire on demographic information. These are described in detail in the following subsection.

3.4 VR-based mental rotation ability test environment

The VMRT used for the pre- and post-measurement of the mental rotation ability was developed and validated in a previous study [11]. The test instrument contained 16 items based on immersive VR and Leap Motion technology. Similar to the paper-and-pencil MRT items [27], the test subjects have to choose two out of four figures corresponding to the target figure. The selection of the figures is possible by grasping them by hand. After the selection, the figures change their colour from blue to green (regardless of whether they are selected correctly or incorrectly) and it is no longer possible to undo this. The given processing time for each task is 30 seconds. After the given time is over, the system automatically goes on to the next task. Finally, the participants are instructed to report to the test supervisor. The instructions are displayed on a board behind the target figures (Figure 2).

3.5 VR-based mental rotation ability training environments

The training of the MRA is enabled by the hands-on interaction with the 3D figures. The persons have the possibility to rotate the figures with their own hands to bring them into a particular position. The goal in the training phase is – similar to the test phase – to find two of four blue figures corresponding to the green figure and to bring them into the same position as the green figure (Figure 2). As soon as the angle of rotation in all three axes (X, Y, and Z axes) is matched (provided the rotated figure was the right choice), the blue figure also changes its colour to green. Maximum deviation tolerance of the turned figure from the target figure is 30 degrees in each axis. If two out of four figures are green, it can be assumed that the task was solved correctly, providing feedback on the task solution. The given time for solving a training task is 40 seconds. After this time, the system automatically switches to the next task. The determination of the next task depends on the solution behaviour of the previous task and the experimental condition. At the beginning of the training, the persons in the adaptive condition are given a task of medium difficulty. If the task is solved in a given time, they will receive more difficult tasks or equally difficult tasks (if the upper difficulty limit is reached). Nevertheless, if the task is not solved, they will receive easier tasks or equally difficult tasks (if the lower difficulty limit is reached). For the difficulty variation of the tasks, the items are selected from four different item pools. The 3D figures in the item pool with the highest level of difficulty consists of six cuboids, while those in the item pool with the lowest level of difficulty are made up of three cuboids. Figure 2, for example, shows the figures with three cuboids.

In the randomised condition, the task is the same as in the adaptive condition. The difference is that in this condition, the tasks from four different item pools are presented randomly, regardless of whether the person has solved the previous task or not.
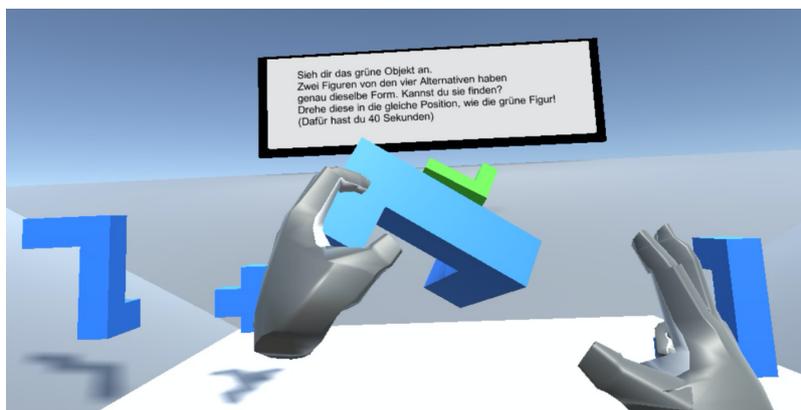

Fig. 2. Scene from the training sequence

3.6 Technology

For the implementation of the software required for the current study, the programme Blender is used to generate the 3D figures, whereas the game engine Unity3D is employed to design the test and training environments.

In order to display the VR-based mental rotation test and training environments, a VR device (HTC VIVE Pro) is connected to a laptop (with a 64-bit Windows 10 operating system, an Intel(R) Core(TM) i7 2.67 GHz processor, 8 GB RAM and an NVIDIA GeForce GTX1070 graphics card) and Leap Motion technology is used. Leap Motion Technology is a small USB device that relies on optical sensors and infrared light to track hands. It is attached to HTC VIVE with a special bracket in the front and middle part of the HMD. Due to this special configuration, it is possible to display hands in VR and to realistically interact with virtual objects.

3.7 NASA-TLX

The NASA task load index (NASA TLX) [56] is used to assess the subjective perceived cognitive load during virtual activity. The authors chose the NASA TLX because it is a standardized multidimensional questionnaire used to measure perceived workload during a task in order to estimate various aspects of performance, including cognitive load. The NASA TLX usually includes six different dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration level. In this experiment, the scale of physical demand is not considered because it is not relevant for the performance of the task. Instead, the participants rate the other dimensions from low to high on a scale of zero to twenty. As expected in the study, the test subjects have to report a lower cognitive load if they receive the tasks that correspond to their respective abilities, and vice versa: The participants will have a higher mental workload if they perform tasks that go beyond their abilities.

3.8 Procedure

The testing and the training take place in a room prepared for the study, where the participants have enough space to move around in the virtual room and view the tasks from different perspectives. First, the test persons fill out a questionnaire on demographic data. After this, they can put on the VR goggles while the test leader makes sure that the HMD is correctly positioned on the head. Before the subjects receive their first VMRT, they complete an exercise sequence consisting of five tasks. This sequence serves to familiarise the subjects with the virtual environment and starts with a scene in which only a single blue figure and the instruction board are visible. The task is to grasp the figure by hand, after which the figure turns green. During this exercise, the test leader gives instructions suggesting to walk around the figure and look at it from all sides.

Further exercises follow that correspond to the tasks of the subsequent VMRT test. In front of the board, a green target object and four different blue answer options are visible. Two identical figures have to be coloured by grasping them by hand. The

participants can mark all four blue pieces green but are not able to undo the action. Therefore, they have to think carefully before deciding which figures they choose to colour. The given time for each exercise is 60 seconds. Then the actual test begins. Each participant has to solve 16 tasks in the test, each within 30 seconds of processing time. At the end of the first phase, a short break is given before the training phase in 3D space follows.

In this phase, every second participant is assigned to an adaptive condition, the other half to the randomised condition. Additionally, in the training phase, there is a first exercise sequence before the training tasks. In the first exercise, the participants have to rotate the single blue figure they are presented with into the same position as the green target object in the background. This type of task is repeated several times so that the test persons are able to familiarise with the handling and rotation of the virtual objects. This is followed by the actual training tasks (20 in total), which, according to the condition, are either presented adaptively or randomly. After the test phase, there is a short break. Then the second VMRT test follows, which is identical to the first test. The whole experiment ends with the NASA-TLX questionnaire.

4 Results

The analysis is carried out with the programme R (version 3.4.4). The packages used are "TAM" [57] for the IRT analysis and "nlme" [58] for the analysis of variance with the linear mixed-effects model. Both of the considered dependent variables, namely the VMRT scores and the NASA TLX data, are normally distributed, allowing the use of methods like IRT analysis, ANOVA (analysis of variance), and t-test.

For the determination of MRA before and after the training, IRT analyses of the solution data are carried out at both times. The analysis is performed according to the criteria of the one-parameter, one-dimensional Rasch model (1PL model) [59]. The mean value of the item difficulties is chosen as the zero point of the logit scale, which represents both item difficulties and personal abilities. The test quality is evaluated by WLE (weighted likelihood estimates) and EAP/PV (expected a posteriori/plausible value) reliabilities [60] as well as by the weighted root mean square values (WMNSQ) within the limits of $0.77 < WMNSQ < 1.33$ [61]. Both IRT analyses of the data show acceptable reliability and fit values. The EAP/PV reliability is .617 (data from the first measuring time) and .746 (data from the second measuring time), respectively, whereas the WLE reliability is .572 and .711, respectively. The weighted mean squares are between 0.849 and 1.273. Personal abilities determined by the IRT analyses are included as dependent variables in further evaluation. The average ability values before and after training, for different conditions and both sexes, are shown in Table 1.

Table 1. Average person abilities in VMRT for different measurement times, conditions, and groups. The abilities are distributed on logit scale and can take values between -3 and +3.

	Mean (M)	Standard deviation (SD)
pre-test	-0.39	0.88
post-test	0.32	1.16
adaptive condition	-0.1	1.12
randomised condition	0.03	1.06
female	-0.28	1.05
male	0.22	1.07

A mixed ANOVA analysis is employed to test the hypotheses by considering several independent variables such as sex (female vs. male), measurement point (before or after training), and condition (adaptive vs. randomised). Subject ID is specified as a random factor. Sex, time of measurement, condition, and their interactions are included as fixed factors.

The 2 group × 2 time × 2 sex mixed ANOVA yielded a main effect of time ($F(1, 98) = 34.343$, $p < .001$) and a main effect of sex ($F(1, 98) = 9.873$, $p < .01$) but no main effect of condition ($F(1, 98) = 0.411$, $p = .41$). The condition × time interaction is also significant ($F(1, 98) = 5.369$, $p < .05$), indicating that MRA improves mor after training in the adaptive condition than it does in the randomised condition. Furthermore, the interactions sex × time ($F(1, 98) = 5.458$, $p < .05$) and sex × time × condition ($F(1, 98) = 6.643$, $p < .05$) are also significant. Figure 3 illustrates these results. It becomes clear that the interactions sex × time and sex × time × condition are significant because only men benefit from the adaptive condition. The possible reasons for this are going to be discussed later. In order to exclude the 3D gaming experience as a confounding factor in the effects shown, correlations between gaming experience and test performances before and after training are determined. The two correlation tests yielded non-significant values ($r1 = .08$, $p = .43$; $r2 = .17$, $p = .08$). Furthermore, the inclusion of the variable 3D gaming experience in the mixed ANOVA did not lead to any significant main effects or interactions, which is why it is excluded from the analysis. Therefore, the 3D gaming experience is not relevant for the VMRT performance and can be neglected as a confounding variable.
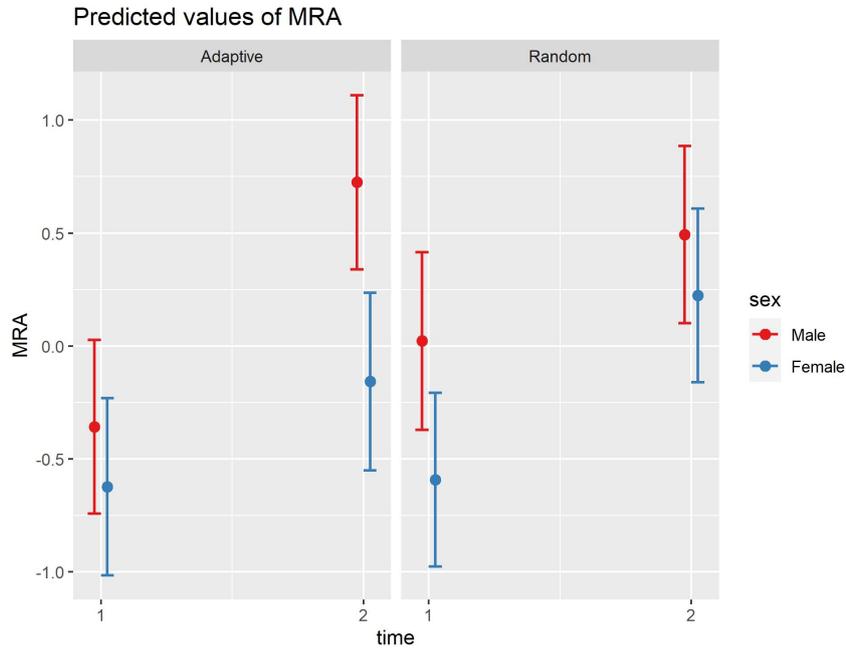
Predicted values of MRA

Fig. 3. Illustration of the results of the mixed ANOVA

In order to investigate whether the cognitive load in the adaptive condition is lower than it is the case in the randomised condition, the mean scale values for the relevant dimensions in the adaptive and randomised conditions are compared using a t-test. The results are shown in Table 2.

Table 2. Comparison of the NASA-TLX scale mean values in the adaptive and randomised conditions. Maximum possible scores on each scale were 20.

Scale	Mean value in the adaptive condition	Mean value in the randomised condition	t-test results
Mental demand	13.902	13.431	$t(100) = 1.651, p = .513$
Temporal demand	10.706	10.863	$t(100) = -0.195, p = .845$
Performance	10.235	11.392	$t(100) =- 1.285, p = .202$
Effort	11.255	11.118	$t(100) = 0.166, p = .869$
Frustration level	9.196	9.843	$t(100) = -0.655, p = .514$

Next, the difference between men and women regarding their cognitive load is examined. In the dimensions mental demand ($t(100) = 1.988, p = .049$) and effort ($t(100) = 2.153, p = .034$), the independent sample t-test shows a lower cognitive load for men than for women. An overview of the results in all dimensions is given in Table 3.

Table 3. Comparison of the NASA-TLX scale mean values for the male and the female test persons. Maximum possible scores on each scale were 20.

Scale	Mean value for the female test persons	Mean value for the male test persons	t-test results
mental demand	14.373	12.961	$t(100) = 1.988, p = .049$
temporal demand	11.373	10.196	$t(100) = 1.481, p = .142$
performance	11.510	10.118	$t(100) = 1.552, p = .124$
effort	12.058	10.313	$t(100) = 2.153, p = .034$
frustration level	10.470	8.568	$t(100) = 1.958, p = .053$

To sum up, the results of the study partially confirm the hypotheses. It could be shown that (1) MRA improves after training, (2) the ability has improved greater in the adaptive condition than in the randomised condition (but only for men), and (3) men generally perform better in VMRT than women. Regarding the cognitive load, there are no differences when comparing the different conditions. The gender differences were found in some dimensions of NASA TLX. The results, as well as weaknesses of the study and further research prospects, are discussed in the next section.

5 Discussion

This experimental pretest-posttest study aimed to find an optimised VR-based training option for MRA. It investigated whether an adaptive environment is able to promote MRA better than a randomised one. The influence of gender and cognitive load during training was considered as well. In the first hypothesis, it was assumed that the training improves MRA, and that therefore, a higher MRA is measurable after the training than before the training. The mixed ANOVA confirmed this hypothesis by indicating that the main effect of time was significant. In addition to the main effect of time, mixed ANOVA also provided a significant main effect of sex, which suggests that men overall perform better in VMRT than women. The gender differences in MRA, which have been proven several times, seem to persist in virtual testing.

The second hypothesis differentiated between the adaptive and the randomised condition and postulated that MRA improvement is greater in the adaptive than in the randomised condition. This hypothesis was also confirmed by the significant interaction condition × time of the mixed ANOVA. However, the triple interaction condition × time × sex showed that the advantage in the adaptive condition is only valid for men. These unexpected results leave several unanswered questions. It is unclear why the adaptive condition might be more beneficial for the male subjects than the randomised condition but shows the opposite for the female participants. From the findings of previous research, the experience with 3D games could be used to explain the advantage of men in the successful use of virtual technology and better results in VMRT [62], [63]. However, investigation of the correlation between 3D gaming experience and VMRT performance yielded no significant results. In our study, this could be attributed to the large majority of subjects having no experience with 3D games (67 out of 102), making comparability with the rest of the sample problematic.

Other factors that may influence VMRT performance could include motivation, technology acceptance, as well as flow or presence experience, which are identified as important factors related to virtual environments. A detailed investigation of this issue would also require a qualitative analysis of training data, which constitutes an interesting topic for future research and should not be ignored.

The third question of the study intended to investigate whether the group differences found by the mixed ANOVA are also reflected in the cognitive load measured by NASA TLX. To be specific, the advantage of the adaptive condition over the randomised condition and the advantage of men over women were to be examined. The t-tests for dependent samples could not detect any difference between the conditions. The gender comparison found a significantly lower effort and mental demand for men than women. The male participants seem to cope more easily with test and training tasks than the female, which can be explained by adducing the higher abilities of the men in mental rotation. However, this effect only proved to be significant in two dimensions of the NASA TLX. In contrast to this. the other three dimensions – temporal demand, performance, and frustration level – were not significant.

At this point, it should be mentioned that NASA TLX, as a subjective measuring method, must be viewed critically because the load can only be assessed retrospectively. For future studies, it would be interesting to include objective measurement methods in the investigation in addition to the subjective measurement methods of cognitive load. As an example, eye tracking or EEG examination could be well combined with the virtual environment (sources) and could provide valuable results regarding MRA.

In addition to NASA TLX, the adaptive training concept described above should be viewed critically as well. The goal was to optimise the virtual training programme by adaptivity, i.e., by tailoring the task difficulties to personal abilities. As mentioned at the beginning of the paper, adaptation can be constructed in many ways. The adaptation described here refers to the adaptation of the task difficulty depending on the solution behaviour of the respective test person. The difficulty of the tasks is exclusively determined by one characteristic, namely the number of cuboids. However, other characteristics could also influence the difficulty of the task. Earlier studies [26] could, for example, identify the rotation angle of the figures to the target figure as a difficulty-determining feature. Furthermore, the adaptive design of the learning environment can consider not only the task character but also individual characteristics of the users. Adapting the material to personal preferences, to motivation, and to the cognitive load of the user could be further steps in designing innovative learning and training environments.

As a follow-up question, the long-term effectiveness of the training seems to be relevant. It would be interesting to investigate whether training effects found with this study could also be measured after days, weeks, or even months. The organisation of the sample in such studies is complex, but it is a necessary prerequisite for the validation of the long-term impact of VR-based MRA training. Furthermore, is should also be questioned to what extent the virtual training of MRA influences the performance in technical areas, or to what extent the training effect can be generalised and applied

to the other domains. Therefore, quasi-experimental studies which examine the training effects in the different MINT subjects would be helpful.

In the future, virtual mental rotation trainings could offer more adaptation possibilities and include more variables than just difficulty adjustments. In addition to mental rotation, personalised training could also be planned for the other dimensions of spatial abilities. One possibility is the personalised training of spatial orientation. This training would require a different environment than rotational training, since for the current study the environment was kept as simple as possible in order to avoid conflicting variables.

In summary, the results of the study partly confirm the expectations. The research question asking whether adaptively designed virtual training can improve MRA can be answered positively, but some questions – and one question in particular – remain: Why do men benefit more from the adaptive training condition than women? In order to investigate this, further studies with additional research objects would be necessary. However, it can be assumed that adaptively designed virtual training environments have great potential for promoting mental rotation. The development and testing of such environments is still in its infancy and requires detailed further investigation.

6 References

[1] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "Computers & Education A systematic review of immersive virtual reality applications for higher education : Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, no. July 2019, p. 103778, 2020, https://doi.org/10.1016/j.compedu.2019.103778.

[2] V. Potkonjak *et al.*, "Virtual laboratories for education in science, technology, and engineering: A review," *Comput. Educ.*, vol. 95, pp. 309–327, Jun. 2016, doi: 10.1016/j.compedu.2016.02.002.

[3] Ü. Çakiroğlu and S. Gökoğlu, "Development of fire safety behavioral skills via virtual reality," *Comput. Educ.*, vol. 133, pp. 56–68, May 2019, https://doi.org/10.1016/j.compedu.2019.01.014.

[4] N. Elmqaddem, "Augmented Reality and Virtual Reality in Education. Myth or Reality?," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 03, pp. 234–242, Feb. 2019, https://doi.org/10.3991/ijet.v14i03.9289.

[5] M. Sattar *et al.*, "Motivating Medical Students Using Virtual Reality Based Education," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 2, pp. 160–174, 2020. https://doi.org/10.3991/ijet.v15i02.11394

[6] C. Landsberg, "Tailoring Instruction To the Individual: Investigating the Utility of Trainee Aptitudes for Use in Adaptive Training," *Electron. Theses Diss.*, 2015, [Online]. Available: http://etd.fcla.edu/CF/CFE0005646/Landsberg_Dissertation_draft_Final.pdf.

[7] D. F. Halpern and M. L. LaMay, "The Smarter Sex: A Critical Review of Sex Differences in Intelligence," *Educ. Psychol. Rev.*, vol. 12, no. 2, pp. 229–246, 2000, doi: 10.1023/A:1009027516424.

[8] J. S. Monahan, M. A. Harke, and J. R. Shelley, "Computerizing the mental rotations test: Are gender differences maintained?," *Behav. Res. Methods*, vol. 40, no. 2, pp. 422–427, May 2008, https://doi.org/10.3758/brm.40.2.422.

[9] D. Voyer, S. Voyer, and M. P. Bryden, "Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables," *Psychol. Bull.*, vol. 117, no. 2, pp. 250–270, 1995, https://doi.org/10.1037/0033-2909.117.2.250.

[10] D. H. Uttal *et al.*, "The malleability of spatial skills: A meta-analysis of training studies.," *Psychol. Bull.*, vol. 139, no. 2, pp. 352–402, May 2013, doi: 10.1037/a0028446.

[11] S. Ariali, "Training of mental rotation ability in virtual spaces," *J. Tech. Educ.*, vol. 8, no. 2, pp. 46–63, 2020.

[12] L. Freina and M. Ott, "A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives," 2015, [Online]. Available: http://www.google.com/patents/US3050870.

[13] T. S. Mujber, T. Szecsi, and M. S. J. Hashmi, "Virtual reality applications in manufacturing process simulation," *J. Mater. Process. Technol.*, vol. 155–156, no. 1–3, pp. 1834–1838, Nov. 2004, https://doi.org/10.1016/j.jmatprotec.2004.04.401.

[14] C. Dede, "Immersive interfaces for engagement and learning," *Science (80-.).*, vol. 323, no. 5910, pp. 66–69, Jan. 2009, https://doi.org/10.1126/science.1167311.

[15] S. Rogers, "Virtual Reality: The learning aid of the 21st century.," *Forbes*, 2019. https://www.forbes.com/sites/solrogers/2019/03/15/virtual-reality-the-learning-aid-of-the-21st-century/#2e403c6a139b.

[16] M. Hoffmann, K. Schuster, D. Schilberg, and S. Jeschke, "Next-Generation Teaching and Learning Using the Virtual Theatre," in *At the Edge of the Rift*, BRILL, 2019, pp. 51–64. https://doi.org/10.1163/9781848883215_006

[17] A. Dünser, K. Steinbügl, H. Kaufmann, and J. Glück, "Virtual and augmented reality as spatial ability training tools," in *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, Apr. 2006, pp. 125–132, https://doi.org/10.1145/1152760.1152776.

[18] C. Carbonell-Carrera and J. L. Saorin, "Virtual learning environments to enhance spatial orientation," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 14, no. 3, pp. 709–719, Nov. 2018, https://doi.org/10.12973/ejmste/79171.

[19] J. Kotlarek, I. C. Lin, and K. L. Ma, "Improving spatial orientation in immersive environments," in *SUI 2018 - Proceedings of the Symposium on Spatial User Interaction*, Oct. 2018, pp. 79–88, https://doi.org/10.1145/3267782.3267792.

[20] A. E. Richardson, D. R. Montello, and M. Hegarty, "Spatial knowledge acquisition from maps and from navigation in real and virtual environments," *Mem. Cogn.*, vol. 27, no. 4, pp. 741–750, 1999, https://doi.org/10.3758/bf03211566.

[21] M. C. Linn and A. C. Petersen, "Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis," *Child Dev.*, vol. 56, no. 6, pp. 1479–1498, Dec. 1985, https://doi.org/10.2307/1130467.

[22] M. Hegarty and D. A. Waller, "Individual Differences in Spatial Abilities," in *The Cambridge Handbook of Visuospatial Thinking*, P. Shah and A. Miyake, Eds. Cambridge University Press, 2005, pp. 121–170. https://doi.org/10.1017/cbo9780511610448.005

[23] R. Colom, M. J. Contreras, J. Botella, and J. Santacreu, "Vehicles of spatial ability," *Pers. Individ. Dif.*, vol. 32, no. 5, pp. 903–912, Apr. 2002, doi: 10.1016/S0191-8869(01)00095-2. https://doi.org/10.1016/s0191-8869(01)00095-2

[24] L. L. Thurstone, *Primary mental abilities.* Chicago: University of Chicago Press, 1938.

[25] J. B. Carroll, *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press, 1993.

[26] R. N. Shepard and J. Metzler, "Mental rotation of three-dimensional objects," *Science (80-.).*, vol. 171, no. 3972, pp. 701–703, Feb. 1971, https://doi.org/10.1126/science.171.3972.701.

[27] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson, "A redrawn vandenberg and kuse mental rotations test - different versions and factors that affect performance," *Brain Cogn.*, vol. 28, no. 1, pp. 39–58, Jun. 1995, https://doi.org/10.1006/brcg.1995.1032.

[28] J. C. Castro-Alonso and D. H. Uttal, "Science Education and Visuospatial Processing," in *Visuospatial Processing for Education in Health and Natural Sciences*, J. C. Castro-Alonso, Ed. Cham: Springer International Publishing, 2019, pp. 53–79. https://doi.org/10.1007/978-3-030-20969-8_3

[29] S. Sorby, B. Casey, N. Veurink, and A. Dulaney, "The role of spatial training in improving spatial and calculus performance in engineering students," *Learn. Individ. Differ.*, vol. 26, pp. 20–29, 2013, https://doi.org/10.1016/j.lindif.2013.03.010.

[30] S. A. Sorby and B. J. Baartmans, "A Course for the Development of 3-D Spatial Visualization Skills," *Eng. Des. Graph. J.*, vol. 60, no. 1, pp. 12–20, 1996, Accessed: Jul. 05, 2020. [Online]. Available: https://eric.ed.gov/?id=EJ528391.

[31] N. L. Veurink and S. A. Sorby, "Longitudinal study of the impact of requiring training for students with initially weak spatial skills," *Eur. J. Eng. Educ.*, vol. 44, no. 1–2, pp. 153–163, Mar. 2019, https://doi.org/10.1080/03043797.2017.1390547.

[32] K. S. Mix *et al.*, "Separate but correlated: The latent structure of space and mathematics across development," *J. Exp. Psychol. Gen.*, vol. 145, no. 9, pp. 1206–1227, Sep. 2016, doi: 10.1037/xge0000182.

[33] F. Sella, E. Sader, S. Lolliot, and R. Cohen Kadosh, "Basic and advanced numerical performances relate to mathematical expertise but are fully mediated by visuospatial skills," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 42, no. 9, pp. 1458–1472, Sep. 2016, https://doi.org/10.1037/xlm0000249.

[34] S. Ariali and B. Zinn, "Virtuelle Umgebungen zur Analyse der mentalen Rotationsfähigkeit," *Band*, vol. 6, no. 4, pp. 7–29, Nov. 2018, Accessed: Jul. 03, 2020. [Online]. Available: http://www.journal-of-technical-education.de.

[35] D. Kimura and E. Hampson, "Cognitive Pattern in Men and Women Is Influenced by Fluctuations in Sex Hormones," *Curr. Dir. Psychol. Sci.*, vol. 3, no. 2, pp. 57–61, 1994, https://doi.org/10.1111/1467-8721.ep10769964.

[36] R. Lynn, "Sex differences in intelligence and brain size: A paradox resolved," *Pers. Individ. Dif.*, vol. 17, no. 2, pp. 257–271, Aug. 1994, https://doi.org/10.1016/0191-8869(94)90030-2.

[37] M. B. Casey, "Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors," *Learn. Individ. Differ.*, vol. 8, no. 1, pp. 39–53, Jan. 1996, https://doi.org/10.1016/s1041-6080(96)90006-0.

[38] M. Kozhevnikov, L. R. Cheng, and M. Kozhevnikov, "Effect of Environment Immersivity on Encoding Strategies of Spatial Tasks," *Procedia Manuf.*, vol. 3, pp. 5059–5066, Jan. 2015, https://doi.org/10.1016/j.promfg.2015.07.519.

[39] R. Seurinck, G. Vingerhoets, F. P. De Lange, and E. Achten, "Does egocentric mental rotation elicit sex differences?," *Neuroimage*, vol. 23, no. 4, pp. 1440–1449, Dec. 2004, https://doi.org/10.1016/j.neuroimage.2004.08.010.

[40] D. M. Adams, A. T. Stull, and M. Hegarty, "Effects of Mental and Manual Rotation Training on Mental and Manual Rotation Performance," *Spat. Cogn. Comput.*, vol. 14, no. 3, pp. 169–198, May 2014, https://doi.org/10.1080/13875868.2014.913050.

[41] G. Wiedenbauer, J. Schmid, and P. Jansen-Osmann, "Manual training of mental rotation," *Eur. J. Cogn. Psychol.*, vol. 19, no. 1, pp. 17–36, Apr. 2007, https://doi.org/10.1080/09541440600709906.

[42] M. Omar *et al.*, "Effects of Mobile Augmented Reality (MAR) towards Students' Visualization," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 20, pp. 106–119, Oct. 2019.

[43] R. Wright, W. L. Thompson, G. Ganis, N. S. Newcombe, and S. M. Kosslyn, "Training generalized spatial skills," *Psychon. Bull. Rev.*, vol. 15, no. 4, pp. 763–771, Aug. 2008, https://doi.org/10.3758/pbr.15.4.763.

[44] A. Wohlschläger and A. Wohlschläger, "Mental and manual rotation," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 24, no. 2, pp. 397–412, 1998, https://doi.org/10.1037/0096-1523.24.2.397.

[45] S. Oxman and W. Wong, "White Paper : Adaptive Learning Systems," *Integr. Educ. Solut.*, 2014.

[46] O. Park and J. Lee, "Adaptive instructional systems," in *Handbook of research for educational communications and technology*, D. Jonassen, Ed. Mahwah, NJ: Lawrence Erlbaum Associates., 2003, pp. 651–685.

[47] C. R. Landsberg, A. D. Mercado, W. L. Van Buskirk, M. Lineberry, and N. Steinhauser, "Evaluation of an adaptive training system for submarine periscope operations," in *Proceedings of the Human Factors and Ergonomics Society*, 2012, pp. 2422–2426, https://doi.org/10.1177/1071181312561493.

[48] K. N. Bauer, R. C. Brusso, and K. A. Orvis, "Using Adaptive Difficulty to Optimize Videogame-Based Training Performance: The Moderating Role of Personality," *Mil. Psychol.*, vol. 24, no. 2, pp. 148–165, Mar. 2012, https://doi.org/10.1080/08995605.2012.672908.

[49] G. Corbalan, L. Kester, and J. J. G. van Merriënboer, "Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement," *Contemp. Educ. Psychol.*, vol. 33, no. 4, pp. 733–756, Oct. 2008, https://doi.org/10.1016/j.cedpsych.2008.02.003.

[50] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Paas, "Cognitive Architecture and Instructional Design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, Jun. 1998, doi: 10.1023/A:1022193728205.

[51] A. Baddeley, "Working memory," *Science (80-.)*., vol. 255, no. 5044, pp. 556–559, 1992, https://doi.org/10.1126/science.1736359.

[52] J. Sweller, "CHAPTER TWO - Cognitive Load Theory," in *Psychology of Learning and Motivation*, vol. 55, J. P. Mestre and B. H. Ross, Eds. Academic Press, 2011, pp. 37–76.

[53] D. Charles *et al.*, "Player-Centred Game Design: Player Modelling and Adaptive Digital Games," in *Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in Play. Digital Games Research Association: DiGRA*, 2005, pp. 285–298. https://doi.org/10.4135/9781446211397.n5

[54] N. Vaughan, B. Gabrys, and V. N. Dubey, "An overview of self-adaptive technologies within virtual reality training," *Comput. Sci. Rev.*, vol. 22, pp. 65–87, Jun. 2016, https://doi.org/10.1016/j.cosrev.2016.09.001.

[55] A. Ewais and O. Troyer, "Authoring Adaptive 3D Virtual Learning Environments," in *Web Design and Development: Concepts, Methodologies, Tools, and Applications*, 2014, pp. 714–733, https://doi.org/10.4018/978-1-4666-8619-9.ch033.

[56] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam: North Holland Press, 1988. https://doi.org/10.1016/s0166-4115(08)62386-9

[57] A. Robitzsch, T. Kiefer, and M. Wu, "TAM: Test Analysis Modules." 2020, [Online]. Available: https://cran.r-project.org/package=TAM.

[58] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, and B. Van Willigen, "Linear and Nonlinear Mixed Effects Models." 2020, Accessed: Jul. 05, 2020. [Online]. Available: https://www.researchgate.net/publication/272475067_The_Nlme_Package_Linear_and_Nonlinear_Mixed_Effects_Models_R_Version_3. https://doi.org/10.1002/9780470087978.ch4

[59] G. Rasch, *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche, 1960.

[60] R. J. Adams, "Reliability as a measurement design effect," *Stud. Educ. Eval.*, vol. 31, no. 2–3, pp. 162–172, Jun. 2005, doi: 10.1016/j.stueduc.2005.05.008.

[61] T. G. Bond and C. M. Fox, *Applying the rasch model: Fundamental measurement in the human sciences: Second edition*. New York: Routledge Taylor & Francis Group, 2007.

[62] M. Murcia-López and A. Steed, "The Effect of Environmental Features, Self-Avatar, and Immersion on Object Location Memory in Virtual Environments," *Front. ICT*, vol. 3, pp. 1–10, Nov. 2016, https://doi.org/10.3389/fict.2016.00024.

[63] R. De Lisi and D. M. Cammarano, "Computer experience and gender differences in undergraduate mental rotation performance," *Comput. Human Behav.*, vol. 12, no. 3, pp. 351–361, Sep. 1996, https://doi.org/10.1016/0747-5632(96)00013-1.

7   Authors

Sunita Ariali is researcher at Department of vocational education focused on teaching technology, University of Stuttgart, Germany. Email: sunita.ariali@ife.uni-stuttgart.de

Bernd Zinn is professor at Department of vocational education focused on teaching technology, University of Stuttgart, Germany.