

Chapter 5

Experiences with Virtual Reality at Secondary Schools: Is There an Impact on Learning Success?



Thomas Keller and Elke Brucker-Kley

5.1 Introduction

Jaron Lanier who is considered a founding father of virtual reality (VR) defines VR among others as an “instrumentation to make your world change into a place where it is easier to learn” (Lanier, 2017). Thus, he describes the potential of immersive VR technologies, which offer 360-degree interactive three-dimensional stimulus environments to engage students in new learning experiences. But, today VR as a promising digital form of experience- and action-oriented learning is still in its infancy in secondary schools in Switzerland. VR learning units for fully immersive VR systems based on head-mounted displays and related to the current learning plans are not commercially available. Usability tests with three prototypically developed VR learning units at secondary schools in Switzerland have shown, however, that the use of VR is considered promising by teachers and that acceptance among students exists (Keller, Hebeisen, & Brucker-Kley, 2018a). However, investments in VR by educational institutions are questionable as long as there is no evidence that the use of immersive VR learning units has an effect on the achievement of learning goals. Beyond providing the basis for convincing business cases for VR-based learning, this research is motivated by the search for design criteria for effective VR learning units.

T. Keller (✉) · E. Brucker-Kley
ZHAW School of Management and Law, Winterthur, Switzerland
e-mail: kell@zhaw.ch; brck@zhaw.ch

© The Author(s), under exclusive license to Springer Nature
Switzerland AG 2021

D. Ifenthaler et al. (eds.), *Balancing the Tension between Digital Technologies
and Learning Sciences*, Cognition and Exploratory Learning in the Digital Age,
https://doi.org/10.1007/978-3-030-65657-7_5

5.2 Research Objective and Approach

The aim of the described research effort is to assess the effect of prototypical VR learning units in mathematics on the learning success of secondary school students. In order to take into account both educational and usability-related criteria for the design and validation of the learning units a three-step research approach is applied and described in the following steps:

1. Prototyping: Design and implementation of a VR learning environment which makes use of the possibilities of immersive VR and implements the requirements of recognized mathematics didactics
2. Testing the usability of the VR learning environment in a controlled experiment with 20 secondary school students (Keller et al., 2018a)
3. Field experiment to evaluate the effect of the VR learning unit on the achievement of predefined learning goals with 67 secondary school students (Keller, Hagen, & Brucker-Kley, 2019)

5.3 State-of-the-Art

5.3.1 Immersive Learning

Learning in virtual worlds is closely linked to the concept of immersion. From a perceptual psychological point of view, immersion describes the feeling of presence in the virtual world, i.e., the illusion of actually being in a virtual reality and being able to interact with it (Cumplings & Bailenson, 2016). Burdea & Coiffet speak (2003) of the so-called three I's—interaction, imagination, and immersion—which must be given in order for a virtual reality to feel real to the user.

From a technical point of view, the following properties must be fulfilled for successful immersion (Slater & Wilbur, 1997):

- If possible, the sensory impressions experienced by humans should be generated exclusively by the computer or one or more output devices
- As many senses as possible should be addressed
- The output devices should completely surround the user
- The output devices should enable a vivid depiction of reality

Currently, this degree of immersion is made possible in commercially available form by so-called Virtual Reality headsets. These VR glasses, which are worn as a head-mounted display (HMD), enable the wearer to move around in a 3D world with a 360-degree view and to experience virtual reality optically and acoustically isolated from the real outside world. This is what distinguishes fully immersive systems from simpler forms of virtual reality such as smartphone-based systems (e.g., Google Cardboard) or desktop-based virtual reality (e.g., Second Life).

Even though innovations in the immersive VR field are still driven by the gaming industry, serious fields of application are becoming increasingly apparent. For

simulations in high-risk training areas (aerospace, medicine, emergency services, workspace safety) as well as for manufacturing and construction productive VR learning environments for vocational training exist (Höntzsch, Katzky, Bredl, Kappe, & Krause, 2013a). For primary, secondary, and higher education two meta-studies by (Mikropoulos & Natsis, 2011) and (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014) conclude that the use of VR can be effective. The effectiveness of immersive 3D learning worlds depends to a large extent on their design being based on both educational and usability-related criteria. (Pirker, Gütl, Belcher, & Bailey, 2013) and (Fowler, 2015) discuss and evaluate models with regard to their applicability to immersive VR learning units. The focus of this valuable research is on the usability of VR learning from the perspective of university students, but not on the effect of VR on the achievement of learning goals by secondary school students. Another research gap arises from the rapidly advancing development of VR technologies and the resulting potential for more effective applications for education. The existing research on learning in virtual worlds and its effects mainly dates from 2006 to 2012 and refers to less immersive desktop-based 3D worlds like Second Life. These learning worlds are only partially comparable to learning units that are implemented for fully immersive VR.

5.3.2 Didactics of Mathematics for Digital Learning Environments

Krauthausen (2012, p. 3) criticizes the fact that computer-assisted learning programs in mathematics that are well-known and widespread on the market today pay too little attention to the current state of research and knowledge in mathematics didactics. The main criticism is that the focus is too much on technology instead of content and the programs thus contribute much to the media competence of learners but little to mathematics skills. He concludes that for the motivation of learners, mathematics itself should increasingly be presented in a way that makes the mathematical content exciting and captivating, rather than packaging the content in an exciting way (Krauthausen, 2012, p. 21). Burrill (2017, p. 316) mentions mathematical accuracy (fidelity) and user experience as central mathematics didactic principles. The mathematical accuracy means that the software should always be mathematically correct, the user experience should not prevent the user from working with the mathematical task and should promote mathematical thinking. Learners need to be able to make decisions to expand their thinking. This possibility is also closely related to the complexity of a task, which does not necessarily require complex mathematical requirements (Geiger, 2017, p. 289). According to Joubert (2017, p. 20), while working on a mathematical task, students apply means from so-called “Modes of Production.” These include acting (usually in the sense of indicating a solution), formulating (developing hypotheses, solution strategies, etc.) and validating (checking based on evidence, theorems, or explanations). Mathematical tasks are intended to encourage learners to do something mathematical and thus experience mathematics in the broadest

sense (Joubert, 2017, p. 4). All tasks should always contain pragmatic and epistemological aspects. The epistemological aspects refer to the insights to be conveyed to learners while working on a task (Sinclair & Zazkis, 2017, p. 177), whereby the pragmatic value of a task is almost always equated with solving the task (Sinclair & Zazkis, 2017, p. 190). Laborde (2011, p. 82) supplements cognitive aspects (what kind of learning the task triggers in the learner at the current state of knowledge), didactic aspects (with what means the task is set) and instrumental aspects (which instruments the learner needs to solve the task).

With regard to pedagogy, Geiger (2017, p. 288) points out in the context of mathematics how eminently important it is to select, adapt, and implement the tasks in the learning environments. In this context, he stresses the importance of cooperation between teachers and researchers in order to anchor well-designed tasks with pedagogically correct approaches in the learning environments. Höntzsch et al. (2013b, p. 4) list the following measures to prevent learners from being overburdened in immersive learning environments:

- clear learning objectives, work orders and instructions,
- permanently available background information,
- hints and exercises that stimulate reflection (for example, setting a specific state of the simulation).

5.4 The VR Learning Units

Design and development of the VR learning environment was based on the ADDIE model according to (Olbrish, 2014). This model consists of five steps: analysis (problem definition), design, development, implementation (application in practice), and evaluation (demonstrating effect).

ADDIE follows game design principles in order to keep the tension of the learner in the so-called flow channel between under- and overchallenge. As recommended in the ADDIE model, the concept was set out in a so-called “Game Design Outline.” This contains, among other things, the objective, storyline, teaching strategy, structure of the game, game components, results (including tracking) and content of the individual scenes (Olbrish, 2014, p. 51). The iterative implementation of the VR prototype was realized with the game engine Unity for a HTC Vive head-mounted display.

To evaluate the artifacts described above, the following design science evaluation methods according to (Hevner, March, Park, & Ram, 2004) were applied:

- Experimental (controlled experiment): usability testing with secondary school students
- Descriptive (substantiated discussion): The artifacts are compared with the state of the art, the requirements and the problem statement.
- In addition, both the iterations of the concept for the learning units and the iterations of the prototype were discussed with teachers in qualitative interviews

Based on the curriculum and the existing maths textbook (Affolter & Walt, 2017), the learning units described in the following section were selected for design and implementation in VR. The selection was driven on the one hand by the suitability of the learning content for immersive learning (imagination, interaction, immersion) and on the other hand by the recommendations and feedback of the involved teachers.

The learning environment is portrayed and can be downloaded as VR app for HTC Vive at www.neulehrkonzepte.ch (Keller, 2017) for HTC Vive.

5.4.1 Unit 1: Introduction to VR

The first virtual learning unit is designed to introduce the students to the immersive experience and the interaction controls. The virtual reality is to be experienced for the first time and the basic interaction concepts for the other learning environments can be experienced and tested. This unit is provided with the intention to prepare the learners so they can fully concentrate on the task in the subsequent learning environments without being distracted by the controls. All learning units are experienced in a single player mode without interacting with other students or teachers. The student learns to move via teleportation, to grab and drop objects and to handle sliders. Duration is 5–10 min.

5.4.2 Unit 2 “Base Area * Height”

The aim of VR unit 2 is to illustrate the connection between the base area, height, and volume of a geometric body and to enable the students to experience their interrelation in three dimensions. Five different geometric bodies must first be placed on their base surface and then a given volume must be achieved by changing the base surface and height in an interactive manner with a slider. Duration is 10–15 min.

5.4.3 Unit 3 “So Big, So Small”

Unit 3 enables the student to interact with hollow masses and lengths in a playful manner. The user has to take different sizes, written as text signs (e.g., 1 L), from a shelf and assign them to a suitable everyday object (e.g., a carton of milk, an aquarium etc.). At the beginning of the unit all objects are displayed in the same size. Only when the solution is requested they take on their correct relative size and the differences become tangible. Duration is 5–15 min.

The next scene works similar, with the difference that linear measures are used (e.g., 1 m). These must be assigned to suitable lengths from everyday life (e.g., the



Fig. 5.1 Sample scene of VR learning unit 3 “So big, so small”

arm span of a person). Thus, length measurements from the tip of a pencil up to 2.5 laps in a sports stadium can be experienced by the student. Duration is 5–10 min (Fig. 5.1).

5.5 Usability Test

Since usability undoubtedly has an influence on the effectiveness of the VR learning units, the prototype was tested in a first phase at four different Swiss secondary schools in the canton of Berne with a total of 20 students. The aim of the usability tests was to identify factors which could possibly favor or hinder the use of the VR learning units with the target group of secondary students. Factors examined included the motivation of the students to learn in such a virtual environment, the learners’ personal feelings about learning success, the personal wellbeing during the experience and the role of gamification elements (rewards). Furthermore, the tests helped to validate the didactic correctness and the quality of the implementation of the learning units.

5.5.1 Sample and Test Procedure

A relevant concern with the use of new technologies in education is that they may not only have potential but also disadvantages for students with special needs. The fact that the VR learning environment described in the previous section was

Table 5.1 Summary of participants of the usability tests

Grade	Age	Gender	Number
7	13–15	Female	1
		Male	7
8	14–16	Female	3
		Male	3
9	15	Female	1
		Male	5
Total number of participating students			20

originally designed for pupils with special needs was a favorable factor for the usability tests. Discussions with the teachers of the participating four schools did show that the idea of inclusion had been implemented for about a year and that students with special needs were integrated into regular classes. In the subject of mathematics, diagnoses for isolated learning disorders (e.g., dyscalculia) were uncommon. Instead, the teachers spoke of partial weaknesses in mathematics. Students with partial weaknesses were integrated into regular mathematics lessons. As a result, the scope for the usability tests was extended. Students with partial weaknesses in mathematics were selected as participants regardless of whether they were diagnosed with a specific disorder. The decision as to who would participate in the evaluation was made by the teachers. Table 5.1 shows the age and number of participants by grade and gender.

The evaluation sessions at each of the four schools lasted between half a day and a full day (depending on the number of participants). If possible, a room was reserved for the entire duration of the evaluation. The test set-up consisted of a powerful PC with sufficient graphical performance, a monitor for observation by the supervisor and the “HTC Vive” system consisting of head-mounted display (HMD, “glasses”), the two controllers and the base stations for motion tracking. The students were taken out of class individually or in pairs and completed the various learning environments in a maximum of 45 min. Before entering the virtual world, the instructions including rules and rights were discussed and the most important operating elements explained. The coaching during the experience in the virtual learning environments was carried out by a researcher co-present as a supervisor outside of the virtual reality environment.

5.5.2 Findings of the Usability Test

Directly after completing the learning units all participants completed an anonymous paper-based questionnaire with 24 questions to evaluate their experience (see Tables 5.2 and 5.3). Questions 1 and 2 ask about the previous experience with virtual reality and video games. Questions 3 to 5 examine the operation and clarity of the tasks. The learning units 2 and 3 are evaluated more precisely with questions 6

Table 5.2 Usability testing—compilation of survey results questions 1–10

#	Age	Class	Sex	Max.	VLU1	VLU2	VLU3	Maximalwert in cm (unterschiedlich auf Grund unterschiedlicher Skalierung verschiedener Drucker)		3.	4.	5.	6.	7.	8.	9.	10.								
								VLU1	VLU2																
1	15	9	male	15.3	X		X	X	14.8	96.7%	15.3	100.0%	0.0	0.0%	14.6	95.4%	15.3	100.0%							
2	14	8	male	15.3	X		X	X	7.6	49.7%	14.9	97.4%	7.5	49.0%	7.7	50.3%	14.9	97.4%							
3	15	9	male	15.3	X	X	X	X	13.9	90.8%	14.9	97.4%	10.2	66.7%	14.7	96.1%	14.2	92.8%							
4	15	9	female	15.1	X	X	X	X	7.4	49.0%	14.3	94.7%	7.2	47.7%	7.5	49.7%	13.2	87.4%							
5	15	8	female	15.1	X	X	X	X	11.1	73.5%	12.9	85.4%	7.6	50.3%	14.2	94.0%	12.4	82.1%							
6	14	8	male	15.1	X	X	X	X	12.7	84.1%	4.1	27.2%	7.5	49.7%	7.7	51.0%	9.7	64.9%							
7	15	9	male	15.1	X	X	X	X	12.6	83.4%	11.3	74.8%	6.5	43.0%	6.4	42.4%	11.4	75.5%							
8	15	9	male	15.1	X	X	X	X	7.5	49.7%	11.6	76.8%	9.7	64.2%	0.0	0.0%	11.3	74.8%							
9	15	9	male	15.1	X		X	X	13.8	91.4%	15.1	100.0%	11.2	74.2%	15.1	100.0%									
10	13	7	female	14.5	X	X	X	X	12.7	87.6%	7.2	49.7%	7.2	49.7%	7.2	49.7%	14.5	100.0%							
11	13	7	male	14.5	X	X	X	X	8.7	60.0%	14.5	100.0%	3.6	24.8%	6.6	45.5%	13.0	89.7%							
12	13	7	male	14.5	X	X	X	X	14.3	98.6%	14.4	99.3%	7.4	51.0%	10.3	71.0%	8.1	55.9%							
13	14	8	female	14.5	X	X	X	X	7.2	49.7%	10.5	72.4%	7.2	49.7%	9.7	66.9%	11.5	79.3%							
14	16	8	male	14.5	X		X	X	6.1	42.1%	7.0	48.3%	6.8	46.9%	7.7	53.1%									
15	15	8	female	14.5	X	X	X	X	10.0	69.0%	13.9	95.9%	7.4	51.0%	10.5	72.4%	8.8	60.7%							
16	14	7	male	14.5	X	X	X	X	11.2	77.2%	13.9	95.9%	10.7	48.8%	11.7	80.7%	3.5	24.1%							
17	13	7	male	14.5	X	X	X	X	8.4	57.9%	7.2	49.7%	7.3	50.3%	13.9	95.9%	11.5	79.3%							
18	15	7	male	14.5	X	X	X	X	11.3	77.9%	11.8	81.4%	8.3	57.2%	14.1	97.2%	13.7	94.5%							
19	13	7	male	14.5	X	X	X	X	13.8	95.2%	14.1	97.2%	9.0	62.1%	9.5	65.5%	13.0	89.7%							
20	13	7	male	14.5	X	X	X	X	14.5	100.0%	9.9	68.3%	9.5	65.5%	11.4	78.6%	14.3	98.6%							
N									20		20		19		20		16		15		15		20		
Max										100.0%		100.0%		74.2%		100.0%		100.0%		66.2%		99.3%		100.0%	
Min										42.1%		27.2%		0.0%		0.0%		55.9%		2.1%		0.0%		44.1%	
Average										74.2%		80.6%		50.2%		64.0%		81.4%		28.7%		61.4%		81.4%	
Median										77.6%		90.1%		50.3%		66.2%		81.4%		25.8%		64.1%		80.8%	
Variance										3.8%		4.8%		2.6%		8.4%		1.8%		3.4%		6.6%		2.4%	
Standard deviation										19.4%		21.9%		16.2%		29.0%		13.4%		18.3%		25.6%		15.4%	

Table 5.3 Usability testing—compilation of survey results questions 11–21

#	Age	Class	Sex	Max.	VLU1	VLU2	VLU3	VLU3	Maximalwert in cm (unterschiedlich auf Grund unterschiedlicher Skalierung verschiedener Drucker)		11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.															
									VLU1	VLU2																										
1	15	9	male	15.3	X		X	X	6.2	40.5%	12.3	80.4%	11.5	75.2%	7.5	49.0%	0.0	0.0%	10.3	67.3%	15.3	100.0%	5.5	35.9%	10.5	68.6%	11.5	75.2%	13.1	85.6%	14.9	97.4%				
2	14	8	male	15.3	X		X	X	6.7	43.8%	7.2	47.1%	12.6	82.4%	11.8	77.1%	0.2	1.3%	12.6	82.4%	14.7	96.1%	0.6	3.9%	7.3	47.7%	4.2	27.5%	14.9	97.4%						
3	15	9	male	15.3	X	X	X	X	2.0	13.1%	12.4	81.0%	12.4	81.0%	15.3	100.0%	0.0	0.0%	15.2	99.3%	12.3	80.4%	1.0	39.6%	15.3	100.0%	11.4	74.5%	15.3	100.0%						
4	15	9	female	15.1	X	X	X	X	7.1	49.0%	12.8	84.8%	11.0	72.8%	13.4	89.7%	4.1	27.2%	12.5	82.6%	12.5	82.6%	5.4	35.8%	13.9	95.4%	10.7	70.9%	11.7	77.3%						
5	15	8	male	15.1	X	X	X	X	6.8	45.0%	1.2	7.9%	11.1	73.5%	14.0	92.7%	5.0	33.1%	11.5	76.2%	12.2	80.8%	12.9	85.4%	11.0	77.5%	14.0	68.9%	12.3	81.5%						
6	14	8	female	15.1	X	X	X	X	3.8	25.2%	11.2	74.2%	5.6	37.1%	13.2	87.4%	1.3	8.6%	11.0	76.2%	12.0	79.5%	12.0	79.5%	10.6	70.2%	7.7	51.0%	14.7	97.4%						
7	15	9	male	15.1	X	X	X	X	2.6	17.2%	0.6	4.0%	14.5	96.0%	7.1	47.0%	6.5	43.0%	7.1	47.0%	13.3	88.1%	11.3	74.8%	13.0	86.1%	10.1	66.9%	13.9	92.1%						
8	15	9	male	15.1	X	X	X	X	2.4	15.9%	15.0	99.3%	11.3	74.8%	15.1	100.0%	10.2	67.9%	0.0	0.0%	14.9	96.7%	3.5	23.2%	13.6	90.1%	7.5	49.7%	11.5	76.2%						
9	15	9	male	15.1	X	X	X	X	2.1	13.9%	13.8	91.4%	14.4	95.0%	14.8	99.0%	0.5	23.2%	4.1	27.2%	12.1	80.1%	6.2	54.3%	12.3	81.5%	3.0	39.9%	13.5	82.6%						
10	13	7	male	14.5	X	X	X	X	6.3	43.4%	6.6	45.5%	7.1	49.0%	10.7	73.8%	0.0	0.0%	14.5	100.0%	11.2	77.2%	11.5	79.3%	14.5	100.0%	10.6	73.1%	14.4	94.9%						
11	13	7	female	14.5	X	X	X	X	6.2	42.8%	0.1	0.7%	7.8	53.8%	10.2	70.3%	7.2	49.7%	14.4	99.3%	14.5	100.0%	5.6	38.6%	12.8	88.3%	6.8	46.9%	14.5	100.0%						
12	13	7	male	14.5	X	X	X	X	3.9	26.9%	10.9	75.2%	14.3	98.6%	10.7	73.8%	0.4	2.8%	13.5	93.1%	14.3	98.6%	7.0	48.3%	14.4	99.3%	8.0	55.2%	14.5	100.0%						
13	14	8	female	14.5	X	X	X	X	6.2	42.8%	6.8	46.9%	10.7	73.8%	9.5	65.3%	6.3	43.4%	12.7	87.8%	6.8	46.9%	12.5	86.2%	9.3	64.1%	9.4	64.8%	14.5	100.0%						
14	16	8	male	14.5	X		X	X	6.2	42.8%	11.3	91.7%	7.4	53.0%	13.3	93.7%	0.4	2.8%	13.3	93.7%	13.1	90.3%	11.3	91.7%	13.6	93.8%	13.1	90.3%	13.1	90.3%						
15	15	8	female	14.5	X	X	X	X	4.9	33.8%	12.5	86.2%	14.5	100.0%	11.4	78.6%	0.0	0.0%	12.2	84.1%	12.5	86.2%	2.5	17.2%	6.9	47.6%	8.4	57.9%	14.5	100.0%						
16	14	7	male	14.5	X	X	X	X	3.9	26.9%			13.7	94.5%	13.8	95.2%	0.5	3.4%	13.5	93.1%	14.0	96.6%	13.8	95.2%	7.2	49.7%	14.1	97.2%	14.0	96.6%						
17	13	7	male	14.5	X	X	X	X	3.5	24.1%	13.5	93.1%	13.6	93.8%	6.7	46.2%	0.2	1.4%	14.3	98.6%	10.8	74.5%	6.9	47.6%	7.0	48.3%	9.5	65.5%	14.3	98.6%						
18	15	7	male	14.5	X	X	X	X	6.1	42.1%	8.4	57.9%	8.4	57.9%	6.7	46.2%	0.2	1.4%	13.0	89.7%	12.1	83.4%	1.0	26.7%	11.2	77.2%	10.2	70.3%	14.0	96.6%						
19	13	7	male	14.5	X	X	X	X	1.3	9.0%	13.2	91.0%	13.7	94.5%	10.1	69.7%	0.3	2.1%	12.4	85.5%	8.5	58.6%	12.8	84.3%	0.4	4.1%	12.3	84.8%	13.4	92.4%						
20	13	7	male	14.5	X	X	X	X	0.1	0.7%	6.6	45.5%	10.0	69.0%	10.2	70.3%	0.3	2.1%	13.8	95.2%	11.6	80.0%	11.4	92.4%	13.9	95.9%	10.8	74.5%	14.4	94.9%						
N									20		19		20		20		20		20		20		20		20		20		20		20		20			
Max										47.0%		99.3%		100.0%		100.0%		67.5%		100.0%		100.0%		95.2%		100.0%		96.2%		100.0%		97.2%		100.0%		
Min										0.7%		0.7%		37.1%		46.2%		0.0%		0.0%		45.9%		3.9%		0.0%		4.1%		19.9%		76.2%		76.2%		
Average										29.8%		63.4%		76.2%		73.1%		15.6%		78.6%		61.9%		59.2%		74.1%		64.1%		74.1%		64.1%		65.5%		93.2%
Median										30.3%		75.2%		75.0%		75.5%		2.8%		86.6%		83.1%		51.3%		79.5%		67.9%		67.9%		67.9%		97.0%		97.0%
Variance										2.1%		10.1%		3.5%		3.4%		4.4%		6.8%		1.9%		9.4%		6.1%		6.1%		3.8%		0.7%		0.7%		
Standard deviation										14.4%		31.7%		18.8%		18.4%		21.0%		26.0%		13.6%		30.7%		24.6%		19.4%		19.4%		14.1%		14.1%		18.1%

to 14 (some questions appear duplicated because they refer to learning unit 2 and 3 separately). The remaining 10 questions deal with the general learning experience.

The answers to question 1 ($N = 20$) show that the majority of the participating students had little experience with VR before the experience (only one person had regular access to a system, eight students had already been in contact or had already seen a VR system).

The frequency with which the test persons play computer games (question 2, $N = 19$) varies greatly. However, no influence on the operation of the learning environments could be observed. Some of those who were very quickly familiar with the operation of the virtual learning environments state that they only very rarely deal with video games. Others, who state that they play a lot, had more trouble with the handling of the VR unit.

In question 5 ($N = 19$) all indicated that they were dependent on the support of the supervisor. Most of them, according to their own statements, needed only a few tips (12) or very little support (5). Two students state that they could not cope with the learning environment without support or at least were heavily dependent on the support of the supervisor.

All of them stated that they liked the two learning units up to perfect (median 81.4% and 80.8%, respectively; question 7 and 10, $N = 16$ and $N = 20$). Almost all of them indicate the level of difficulty (questions 8 and 11) between easy and medium. Only two describe the learning units as rather difficult ($N = 15$ and $N = 20$ respectively). The observations did show that few students were able to solve the tasks (especially learning unit 3) directly without having to think and correct their first solution. For learners who have succeeded in doing so, there is a small tendency for them to assess their concentration. (Question 16) and their learning success (Question 20) lower.

The place value chart (question 12, $N = 19$) implemented as a shelf in virtual learning unit 3, a didactic element of the known math textbook, was not recognized by 3 learners according to their statements. The others had already seen it, with a tendency to know it well. This didactic material is therefore also well recognized in its virtual form of presentation. Only 9 immediately recognized the everyday objects (question 13, $N = 20$). The others did not recognize all of them immediately. According to the observations, this particularly affects the 3D model of a child's arm span (1 m, often thought to be a doll not a person) and the syringe and ink cartridge of the hollow masses (even in original size the difference in size was often not recognized). Scales had been added as supporting aids during the implementation.

Seeing the objects in their original size has helped all participating students (mostly a lot) (question 14, $N = 20$). It therefore seems to have succeeded in making it possible to experience orders of magnitude that are difficult to comprehend on paper and to point out errors simply, comprehensibly, and impressively.

According to the learners (question 16, $N = 20$), learning in the virtual learning environment has a positive influence on concentration. Three learners state to have been more concentrated than in class, 14 students state that they have even been

very concentrated. The 3 test participants (all from the ninth grade), who stated lower values, had the subjective impression that the tasks were rather too easy for them. This is also confirmed by their answers on the level of difficulty and subjectively perceived learning success.

What is surprising is the effect of the trophies which were implemented as a minimal form of gamification to increase the motivation of the learners. This was observed during the evaluation and is also clearly reflected in the questionnaire in the answers to question 19 ($N = 19$) (median 79.5%). Only one person states that the awards hardly motivated him. 10, on the other hand, have been very motivated, 5 also indicate a strong positive influence on motivation and the remaining 4 have at least been somewhat motivated.

With an overwhelmingly high value, all participants state that they would very much like to have lessons in virtual reality again (question 21, $N = 20$, median 97.0%, minimum 76.2%). The positive effect of the novel medium on motivation seems to be given. The fact that half of all learners say that learning in the virtual environment felt more than half (or even completely) like school (question 18, $N = 20$), and that everyone claims to have learned something (mostly much, median 67.9%) (question 20, $N = 20$) also points out that the interest in learning with VR is not just an escape from regular school.

5.6 Field Experiment

5.6.1 *Design of the Field Experiment*

In the field experiment a slightly modified version of the pre-test/post-test control group design according to (Campbell & Stanley, 1967) was applied. A written pre-test was conducted and graded at the beginning of the field experiment to assess the level of knowledge in relation to the learning goals covered by the learning unit (see Sect. 5.6.2). Then the independent variable, i.e. the exposure to the VR learning environment, was changed, and a post-test was performed. In addition to this second test, a third test was performed 1 month later to measure medium-term learning success. The external validity of the experiment was increased by not informing the students that they are part of a field experiment.

Four classes of a Swiss secondary school in the canton of Zurich served as comparison groups. This school was not part of the usability tests described in Sect. 5.5. In the public school system of the respective canton, secondary school classes are divided into three categories—Sek A, Sek B, Sek C—based on cognitive requirements, with A being the most demanding category. Two of the four classes participating in the field experiment belong to the higher performance level A and two classes to the lower performance level B. One of the two classes of each category was in the experimental group (VR unit) and the other one in the control group (regular unit). In total, the comparison groups included 87 students, 45 of whom

belonged to a Sek A class and 42 to a Sek B class. Due to drop-outs in course of the 3 tests of the field experiment, 67 subjects could be included in the final analysis. Of the 67 valid subjects 34 were in the experimental group and experienced the VR unit, while the other 33 subjects were in the control group and attended a regular mathematics lesson targeted at the very same predefined learning goals (see Sect. 5.6.2).

The randomization in this field experiment was given by the class distribution of the school. Thus, the field experiment must be regarded as quasi-randomized.

A protocol was kept during the execution of the field experiment, which recorded exceptional events such as assistance or technical problems. Furthermore, the students of the experimental group assessed their subjective learning success and learning experience in a verbal interview based on a structured questionnaire directly after experiencing the VR learning units.

5.6.2 Refinement of Learning Goals

In order to ensure that the VR lesson and the conventional mathematics lesson have the same learning objectives and to check the level of knowledge in the pre-test and the two post-tests in a comparable way, verifiable learning goals had to be specified starting from the learning goals driving the design of the VR learning units. Ten cognitive learning goals on all six layers of the Bloom taxonomy of learning goals were defined in close collaboration with the involved teachers (Bloom, Krathwohl, & Masia, 1984):

1. Knowledge “involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting.”
2. Comprehension “refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications.”
3. Application refers to the “use of abstractions in particular and concrete situations.”
4. Analysis represents the “breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between ideas expressed are made explicit.”
5. Synthesis involves the “putting together of elements and parts so as to form a whole.”
6. Evaluation engenders “judgments about the value of material and methods for given purposes.”

This resulted in the following learning goals matched to the six taxonomy levels according Table 5.4.

Table 5.4 Goal—taxonomy level mapping

Learning goal	Taxonomy level
The student knows the different units of hollow measures and can arrange them according to size	1 and 2
The student knows the different units of measures of length and can arrange them according to size	1 and 2
The student is able to determine/imagine the hollow measures of everyday objects without using any aids	3
The student is able to determine/imagine the measures of length of everyday objects without using any aids	3
The student is able to compare the hollow measures with each other and put them in relation to each other without using any aids	4
The student is able to compare and relate length measures to each other without using any aids	4
The student is able to assign hollow measures to everyday activities, for example, how much water she or he drinks per day	5
The student is able to assign length measures of lengths to everyday activities, for example, the length of the way to school	5
The student can check the correctness of a statement or task regarding hollow measures without using any aids	6
The student can check the correctness of a statement or task regarding measures of length without using any aids	6

5.6.3 Learning Unit Without VR

The conventional lesson without the use of VR for the control group was performed by the teachers themselves in order to maintain the authenticity of the field experiment. The content of the lesson and the paper-based exercise were based on the same math textbook chapters as the VR learning unit and were geared toward the same learning goals as listed above.

5.6.4 Evaluation of the Results of the Field Experiment

Both the immersive VR learning experience and the conventional teaching lesson have resulted in short and medium-term learning success for the students in both comparison groups.

The subjects in the experimental group (with VR) between the pre-test (test 1) and the first post-test (test 2) achieved an average learning success of 1.38 points (19.14%). In the control group, the average learning success between the same tests was 1.30 points (18.49%). The medium-term learning success, which is measured by the difference between test 1 and the second post-test (test 3), averages 0.64 points (8.88%) for the subjects in the experimental group. The subjects in the control group recorded an average learning success of 0.45 points (6.40%) between the same tests (Fig. 5.2).

AVERAGE TEST RESULTS

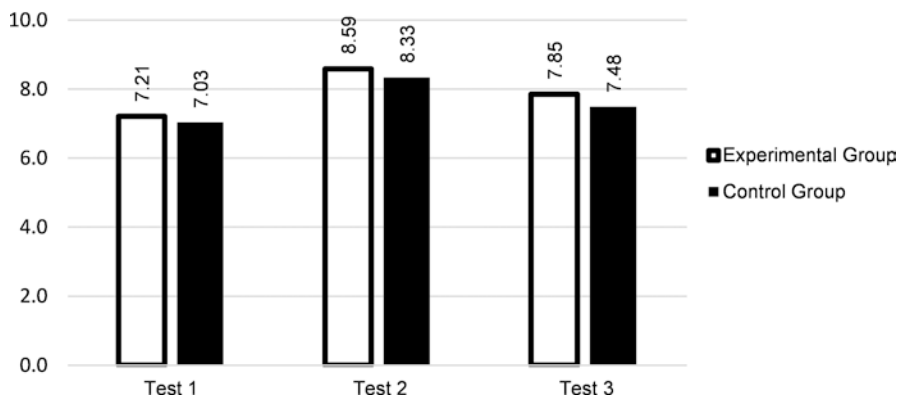


Fig. 5.2 Average values of test results of experimental and control group

Table 5.5 Comparison between experimental with control group

Change in learning success	Experimental group	Control group	Difference	Statistical significance
From test 1 to test 2	1.38	1.30	0.08	58.24%
From test 1 to test 3	0.64	0.45	0.19	64.71%

If the average test results of the subjects in the experimental and control groups are compared, the difference between test 1 and test 2 is 0.08 points. The statistical significance is 58.24%, which means that the effects of the two teaching methods do not differ. Similar observations can be made for the measured average learning outcomes between test 1 and test 3. There it is a difference of 0.19 points with a statistical significance of 64.71% (Table 5.5).

In summary, this means that both immersive VR and the conventional teaching lesson have resulted in statistically verifiable learning success for the test persons. If the learning outcomes of the different groups are compared, however, there are no differences. Both ways of teaching are to be regarded as equivalent in terms of both short-term and medium-term learning success in this setting. However, there are tendencies, if the questions of the random sample tests are considered individually, that the tasks with reference to the measures of length were better solved by the test persons of the experimental group and the tasks with reference to the hollow masses were better solved by the test persons of the control group.

In order to be able to make more precise statements about the learning success of the test persons, the test results of the comparison groups were analyzed by the performance level of students given by their affiliation to the categories of Sek A (high) and Sek B (lower). Their average test results show that there are no striking differences in the achievement of measurable learning goals of the test subjects between test 1 and test 2 (Fig. 5.3). However, if one looks at test 3, it is noticeable

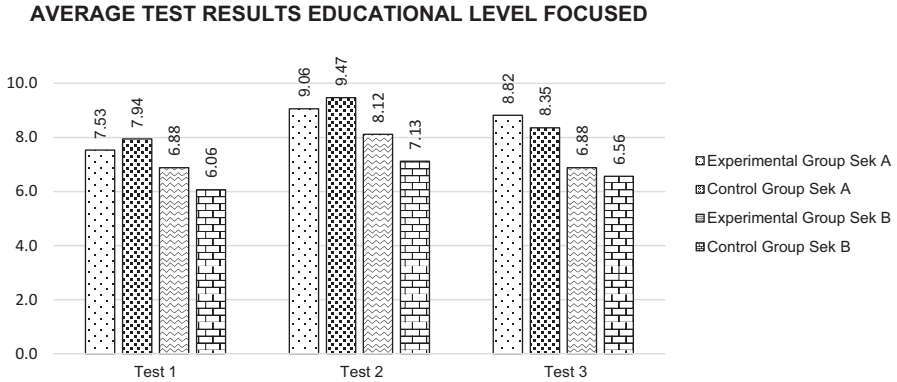


Fig. 5.3 Average test results by performance levels

Table 5.6 Learning success between test 1 and test 2

Comparison of subjects of Sek A	Experimental group	Control group	Difference	Statistical significance
Learning success test 1 to test 3	1.24	0.41	0.83	88.50%

Table 5.7 Comparison of learning success between levels

Comparison of experimental group	Subjects of Sek A	Subjects of Sek B	Difference	Statistical significance
Learning success test 1 to test 3	1.24	0.00	1.24	99.01%

that the test subjects in Test Group Sec A achieved the highest average score, which was not the case in the previous two tests.

The average learning success between test 1 and test 3 was 1.53 points (16.33%) for the subjects in the Sek A test group. The average learning success achieved by the test subjects in the Sek A control group was 0.41 points (5.16%), which is considerably lower.

If the average test results between test 1 and test 3 of the subjects in the Sek A category of both comparison groups are compared, the difference between the average learning successes is 0.83 points (Table 5.6). The statistical significance is 88.50%. This means that the difference is not statistically relevant, but there is a tendency that should be further investigated.

If the average learning success of the subjects in Sek A is compared with that of the subjects in Sek B of the experimental group between test 1 and test 3, this shows a difference of 1.24 points. In addition to the comparatively high learning success of the subjects in Sek A, this is mainly due to the fact that the subjects in Sek B did not achieve any learning success in these tests. With a statistical significance of 99.01%, this difference is considered statistically relevant (Table 5.7).

In summary, this means that immersive teaching methods led to verifiable short and medium-term learning success for subjects in Sek A, in contrast to the students

AVERAGE TEST RESULTS GENDER FOCUSED

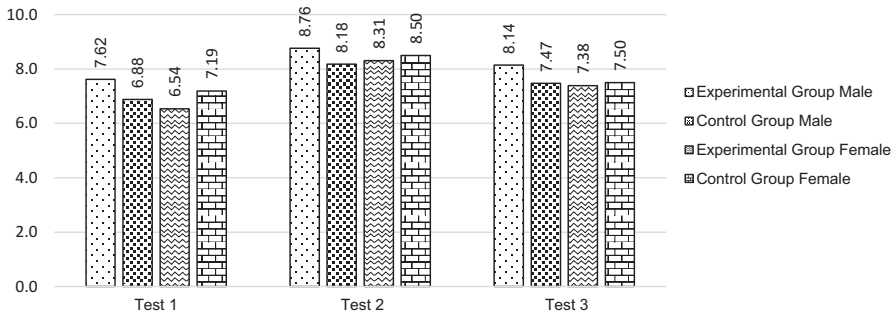


Fig. 5.4 Average test results by gender

QUESTION 1A GENDER FOCUSED

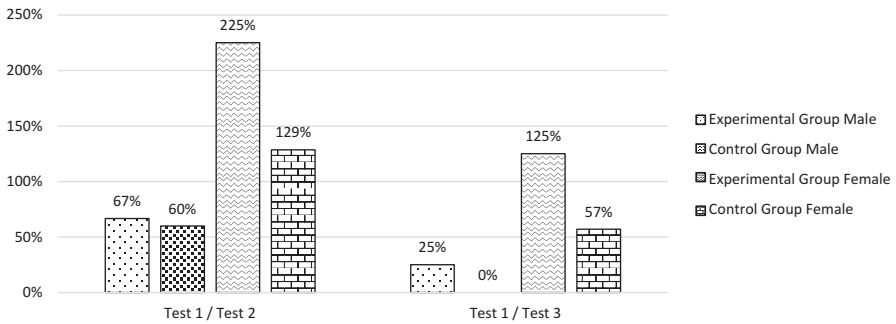


Fig. 5.5 Question 1A visualized by gender

in Sek B, who only achieved short-term learning success. Consequently, there is a difference between the two performance categories levels with regard to medium-term learning success with VR learning units.

In a next step the test results of both control and experimental groups were analyzed by gender. As can be seen in Fig. 5.4, the average test results are similar. The female students in the experimental group who achieved the highest learning success between test 1 and test 2 with 1.77 points (27.06%) are notable. This finding is particularly evident when the focus is placed on the individual questions of the random sample tests. For example, the female respondents in the experimental group in question 1a, which related to the measures of length in the learning unit, have by far the highest average learning success with 225.00% between test 1 and test 2.

As shown in Fig. 5.5, the female respondents in the control group are far below this with an average learning success of 128.57%. The male subjects in the experimental group were also lower with an average learning success of 66.67%. Similar

observations can be made for the learning success between test 1 and test 3. With 125.00% learning success, the female subjects in the experimental group are far ahead of the female subjects in the control group with 57.14% and the male subjects in the experimental group with 66.67%.

If, on the other hand, the average learning success between test 1 and test 2 of the female subjects in the test and control groups are compared, the difference is 0.46 points, with the test group scoring higher. The statistical significance is 78.39%. A similar observation can be made between test 1 and test 3, where the difference is also 0.46 points. Here, the statistical significance is somewhat lower at 74.82%. This means that both learning units are of equal value for the female subjects.

If the average learning successes of the male and female subjects in the comparison group are compared, there is a difference of 0.63 points, with the female subjects being higher.

In summary, this means that there are no differences between the subjects in the experimental and control groups when analyzed by gender. The effects of the VR and the conventional learning units are to be considered as equivalent. In addition, there are no differences between male and female subjects who used the VR unit. However, there is a tendency that the female subjects benefited more from the VR units than the male subjects.

5.6.5 Observations and Feedback of Students

The positive effect on motivation and possible new experiences mentioned in the literature was confirmed by the observations and the verbal survey among the students experiencing the VR learning units. Almost all students have worked in a very concentrated manner and state this in the questionnaire. The potential for addiction quickly became apparent in course of the experiment. Questions were asked about the possible use of VR technology for video games and many stated in the answers to the questionnaire that they already spend a lot of time with video games every day. A constructive discussion with the learners about addictive media behavior and content in private use appears to be sensible and, alongside clear rules of conduct seems relevant to counteract problematic media behavior.

The need for coaching, help, and feedback in the VR learning units also became evident. Most learners say that they were dependent on the support of the physically present supervisor and were able to concentrate on them. These feedbacks indicate, that most of the subjects could be kept in the flow channel (Olbrish, 2014, see Sect. 5.4) during the evaluation, as the observations of the supervisor during the experiences also confirm. The successful implementation of the ADDIE model could also be concluded from the minor correlation that has been recognized in the answers on task difficulty, concentration, and learning success. The positive effect on self-assessed learning success and motivation mentioned in the literature can be confirmed. With a few exceptions, all learners indicated a positive learning success in the responses to the questionnaire. Many learners were able to observe how they could expand their personal experience space (eureka moment). However, empirical proof is still lacking.

5.7 Conclusions and Outlook

The analysis of the field experiment showed that both learning units resulted in short and medium-term learning success for the test persons, but that there were no differences between the effects of the two teaching approaches. A difference can be observed between the educational levels. In contrast to the students in Sek B, the students in Sek A had a better learning success in the medium term.

Statistically, there are no significant differences by gender. However, there are indications that female subjects benefit more than male subjects. This tendency could not be statistically proven in the context of the present study. It concerns both short and medium-term learning success. This is a finding that could be further investigated within the framework of future research.

The effectiveness of the tested VR learning units differs by content. Measures of length performed better than hollow measures. It remains to be examined whether this really has to do with the specific content or with the specific design of the respective learning units.

Furthermore, an important limitation must be mentioned. There is no doubt that the novelty of VR has an influence on the test persons. This can have a positive effect (motivation), but also a negative one (distraction). In future field experiments this can only be addressed if the students work with VR learning units or other VR applications for a longer time in advance that have no relation to the content of tested learning units.

Finally, one must conclude once more that the use of VR learning units does not per se lead to better learning success (Keller et al., 2019; Keller, Hebeisen, & Brucker-Kley, 2018b). As with conventional teaching methods, the quality and design of such VR learning units play a decisive role. Exploratory experiments that explore the influence of known and new design criteria on different learning outcomes beyond measurable test results are considered very relevant. In particular, the design and the effect of a help and feedback system for VR learning environments appeared to be a promising research topic in both the usability test and the field experiment.

References

- Affolter, W., & Walt, M. (2017). mathbuch IF: Begleitband für die integrative Förderung Klassen 7–9 (1. Auflage). Klett und Balmer Verlag.
- Bloom, B. S., Krathwohl, D. R., & Masia, B. S. (1984). *Taxonomy of educational objectives. The classification of educational goals: Cognitive domain handbook 1*. New York: Longman.
- Burdea, G., & Coiffet, P. (2003). *Virtual reality technology* (2nd ed.). Hoboken, NJ: J. Wiley-Interscience.
- Burrill, G. (2017). Designing interactive dynamic technology activities to support the development of conceptual understanding. In A. Leung & A. Baccaglioni-Frank (Eds.), *Digital technologies in designing mathematics education tasks: Potential and pitfalls* (Vol. 8, pp. 303–328). Basel: Springer.
- Campbell, D. T., & Stanley, J. C. (1967). *Experimental and quasi-experimental designs for research (2. Print)*. Boston: Houghton Mifflin Comp.

- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412–422. <https://doi.org/10.1111/bjet.12135>
- Geiger, V. (2017). Designing for mathematical applications and modelling tasks in technology rich environments. In A. Leung & A. Baccaglioni-Frank (Eds.), *Digital technologies in designing mathematics education tasks: Potential and pitfalls* (Vol. 8, pp. 285–302). Basel: Springer.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*, 75–110.
- Höntzsch, S., Katzky, U., Bredl, K., Kappe, F., & Krause, D. (2013a). Simulationen und simulierte Welten. In *Lehrbuch für Lernen und Lehren mit Technologien: 2. Auflage*, p. 327.
- Höntzsch, S., Katzky, U., Bredl, K., Kappe, F., & Krause, D. (2013b, January 1). Simulationen und simulierte Welten: Lernen in immersiven Lernumgebungen [InternetDocument]. Retrieved from <http://13t.tugraz.at/index.php/LehrbuchEbner10/article/download/102/108>
- Joubert, M. (2017). Revisiting theory for the design of tasks: Special considerations for digital environments. In A. Leung & A. Baccaglioni-Frank (Eds.), *Digital technologies in designing mathematics education tasks: Potential and pitfalls* (Vol. 8, pp. 17–40). Basel: Springer.
- Keller, T. (2017). Neue Lehrkonzepte an Schulen – Einsatz von Virtual Reality. Retrieved February 25, 2020, from <http://neuelehrkonzepte.ch/>
- Keller, T., Hagen, F., & Brucker-Kley, E. (2019). A field study about the impact of a VR learning unit. In *CELDA 2019 16th International Conference Cognition and Exploratory Learning in Digital Age, Cagliari, Italy, 7–9 November 2019* (pp. 307–314). IADIS Press.
- Keller, T., Hebeisen, A., & Brucker-Kley, E. (2018a). Integration of children with special needs in mathematics through virtual reality, pp. 30–37. Retrieved from <http://www.iadisportal.org/digital-library/integration-of-children-with-special-needs-in-mathematics-through-virtual-reality>
- Keller, T., Hebeisen, A., & Brucker-Kley, E. (2018b). Integration of children with special needs in mathematics through virtual reality, pp. 30–37. Retrieved from <https://digitalcollection.zhaw.ch/handle/11475/16019>
- Krauthausen, G. (2012). Digitale Medien im Mathematikunterricht der Grundschule. <https://doi.org/10.1007/978-3-8274-2277-4>
- Laborde, C. (2011). Designing substantial tasks to utilize ICT in mathematics lessons. In A. Oldknow & C. Knights (Eds.), *Mathematics education with digital technology* (pp. 75–83). New York: Bloomsbury. <https://doi.org/10.5040/9781472553119>
- Lanier, J. (2017). Dawn of the new everything: Encounters with reality and virtual reality. Retrieved from <https://us.macmillan.com/dawnoftheneweverything/jaronlanier/9781627794091/>
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. <https://doi.org/10.1016/j.compedu.2013.07.033>
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769–780. <https://doi.org/10.1016/j.compedu.2010.10.020>
- Olbrish, K. (2014). *Immersive learning: Designing for authentic practice*. Alexandria, VA: ASTD Press.
- Pirker, J., Gütl, C., Belcher, J. W., & Bailey, P. H. (2013). Design and evaluation of a learner-centric immersive virtual learning environment for physics education. In A. Holzinger, M. Ziefle, M. Hitz, & M. Debevc (Eds.), *Human factors in computing and informatics* (pp. 551–561). Berlin: Springer.
- Sinclair, N., & Zazkis, R. (2017). Everybody counts: Designing tasks for TouchCounts. In A. Leung & A. Baccaglioni-Frank (Eds.), *Digital technologies in designing mathematics education tasks: Potential and pitfalls* (Vol. 8, pp. 175–192). Basel: Springer.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments. *Presence: Teleoper. Virtual Environment, 6*(6), 603–616. <https://doi.org/10.1162/pres.1997.6.6.603>