UNIVERSITY OF TURKU

Interaction design, Virtual reality and Education: Principles of Designing Virtual Reality Learning Materials –

Use Case: Biology Class in Finnish Lower Secondary School

Interaction Design
Master’s Degree Program in Information and Communication Technology
Department of Computing, Faculty of Technology
Master of Science Thesis

Author:
Maiju Kyyhkynen

Supervisors:
D.Sc. (Tech.) Teijo Lehtonen (University of Turku)
PhD Taneli Nyyssönen (University of Turku)

December 2025

Master’s thesis

Subject: Interaction Design
Program: Master’s Degree Program in Information and Communication Technology
Author: Maiju Kyyhkynen
Title: Interaction design, Virtual reality and Education: Principles of Designing Virtual Reality Learning Materials – Use Case: Biology Class in Finnish Lower Secondary School
Number of pages: 111 pages, 2 appendix pages
Date: December 2025

Virtual reality (VR) is a novel technology in the field of education and found promising regarding its ability to motivate students and to enhance learning. NASA and the military developed VR devices originally for simulation training purposes and since then a breakthrough in affordable custom products for the general public has been expected to happen. The technology has improved from the early models, and today the capacity and the quality of devices are getting close to the level of comfortable use. However, adopting this new technology into education requires research as well as quality learning materials which are still both found lacking.

The aim of this study is to recognize elements of valuable VR learning material, features to include and elements that can challenge or hinder learning. The study combines principles of interaction design, education, and VR to build a framework for design purposes. The first component, interaction design is to ensure the comfortable use of material and to guarantee that essential information is provided for users. The second component, VR is the technology in interest, and the third component, education sets environmental boundaries for the design providing the pedagogical aspect and defining the learning approach. The four phases of the study comprehend the teachers’ questionnaire (n=18), the development of the application, the students’ user testing and feedback, and the results. The first phase is to find the features that will be included in the VR application designed. The national curriculum and previous research findings are utilized in the structuring and formulation of questionnaires and in the design of the test setting. After building the VR application which follows the approach of inquiry learning, the material is tested by students of lower secondary school (n=17) in an authentic classroom context in the third phase of the study. After the testing, students are asked to give feedback (n=14) and to fill out a questionnaire. In the final phase, the data is analyzed, and the results are discussed.

The study reveals that VR is seen advantageous in the field of education both by teachers and students and many possible topics for VR sessions are suggested. In VR learning, most of the twelve quality requirements found relate to themes of usefulness, and students’ experience, but usability and safety aspects are seen important as well. Teachers with over ten years of teaching experience expect significantly more usable learning material and prefer to have a considerably less active role for themselves during VR sessions. The user testing of the learning material reveals being mainly a positive experience for students, but some inadequacies have risen from the material that relate to usability and safety themes: less discomfort (e.g., nausea, headache, migraine) and more comprehensive instructions are needed to meet students’ expectations. These findings suggest that the functionality of VR is not yet fully understood and more research is needed. Especially if subject-related learning is the aim of the study, it is important to pay attention to the sufficient sample size, the test setting and the fact that there is enough time and VR headsets available. Functional learning materials (i.e., tools and platforms) are proposed as the focus of design to enable teachers and students to make the material themselves. Participatory design approach revealed to be valuable in designing and finding challenges of VR learning. The framework used can be recommended for further studies and as a tool to develop better VR learning materials in the future.

Key words: interaction design, virtual reality, education, learning material, inquiry learning

Pro gradu -tutkielma

Linja: Vuorovaikutusmuotoilu
Ohjelma: Tietojenkäsittelytieteiden maisteriohjelma
Tekijä: Maiju Kyyhkynen
Otsikko: Vuorovaikutusmuotoilu, virtuaalitodellisuus ja opetus: Virtuaalitodellisuusoppimateriaalisuunnittelun periaatteet – Käyttötapaus: biologian tunti suomalaisessa yläkoulussa
Sivumäärä: 111 sivua, 2 liite sivua
Päiväys: Joulukuu 2025

Virtuaalitodellisuus (VR) on opetusalalle uusi ja lupaava teknologia, joka motivoi oppilaita ja parantaa oppimista. NASA ja armeija kehittivät VR-laitteita alun perin simulaatiokoulutustarkoituksiin ja tavallisille ihmisille sopivia, huokeita VR-laitteita ja teknologian läpimurtoa on odotettu kauan. Tekniikka on parantunut ensimmäisistä malleista ja nykyään laitteiden suorituskyky ja ominaisuudet ovat jo hyvin lähellä mukavaa käytön edellyttämää rajaa. Tarvitaan kuitenkin tutkimusta ja laadukasta oppimateriaalia, ennen kuin uusi teknologia voidaan omaksua opetukseen. Molemmista edellä mainituista löytyy vielä puutteita ja parannettavaa.

Tämän tutkimuksen tarkoituksena on tunnistaa ominaisuuksia, joita hyvän VR-oppimateriaalin tulisi sisältää ja toisaalta löytää ominaisuudet, jotka voivat haastaa tai estää oppimista. Tutkimus yhdistää vuorovaikutusmuotoilun, opetuksen ja VR:n periaatteita suunnittelumallin rakentamiseksi. Ensimmäisen komponentin, vuorovaikutusmuotoilun, tarkoituksena on varmistaa oppimateriaalin käyttömukavuus ja varmistua, että kaikki tarpeellinen tieto on käyttäjän saatavilla. Toinen komponentti, VR, on teknologia, jota tutkimuksessa hyödynnetään. Kolmas ja viimeinen komponentti on koulutus, joka toimii ympäristönä ja asettaa rajat suunnittelulle tarjoten mm. pedagogisen näkökulman ja oppimisteorian. Tutkimuksen neljä vaihetta sisältävät opettajille tehtävän kyselyn (n = 18), VR sovelluksen rakentamisen, oppilaiden käyttäjätestauksen ja palautteen keräämisen ja tulosten käsittelyn. Ensimmäisessä vaiheessa selvitetään mitä ominaisuuksia VR sovelluksen pitäisi sisältää. Opetussuunnitelmaa ja aikaisempia tutkimustuloksia aiheesta, käytetään apuna kyselyiden rakentamisessa ja muotoilussa sekä koeasetelman suunnittelussa. Kolmannessa vaiheessa tehdään VR-sovellus, joka seuraa tutkivan oppimisen periaatteita. Sovellusta testataan yläkoulun oppilailla (n = 17) todellisessa luokkatilanteessa. Testauksen jälkeen oppilaita pyydetään antamaan palautetta (n=14) täyttämällä kyselylomake. Viimeisessä vaiheessa kerätty aineisto analysoidaan ja tulokset käydään läpi.

Tutkimuksessa käy ilmi, että sekä opettajat, että oppilaat näkevät VR:n opetuksessa hyödylliseksi teknologiaksi ja mainitsevat monia mahdollisia aiheita VR tunneille. Suurin osa opettajien asettamista laatuvaatimuksista liittyy hyödyllisyysteemaan ja oppilaan kokemukseen, mutta myös käytettävyys ja turvallisuusnäkökulmat nähtiin tärkeiksi. Opettajat, joilla on yli kymmenen vuoden opetuskokemus, odottavat merkitsevästi parempaa käytettävyyttä materiaaleilta ja toivovat itselleen vähemmän aktiivista roolia VR-oppimisessa. VR-oppimateriaalin käyttäjätestaus osoittautui oppilaille pääasiassa positiiviseksi kokemukseksi, mutta oppilaat löysivät myös puutteita liittyen käytettävyyteen ja turvallisuuteen. He toivoivat, että epämukavuuden kokemus vähenisi (esim. huonovointisuus, päänsärky, migreeni) ja, että parempia ohjeita olisi saatavilla. Saadut tulokset osoittavat, että VR:n toimintaperiaatteita ei vielä täysin ymmärretä ja lisää tutkimusta tarvitaan. Erityisesti, jos halutaan tutkia aineenoppimista, on tärkeää huomioida riittävä otoskoko, koeasetelma ja varata tarpeeksi aikaa ja VR-laitteita käyttöön. Suunnittelun painopisteen tulisi olla ensi sijassa oppimateriaalityökalujen ja oppimisalustojen suunnittelussa, joiden avulla opettajat ja oppilaat voivat itse tehdä erilaisia materiaaleja. Osallistava suunnittelu osoittautui hyväksi tavaksi suunnitella ja löytää VR-oppimisen haasteita. Tutkimuksessa käytettyä mallia voi suositella käytettäväksi myöhemmissä tutkimuksissa ja jatkossa myös työvälineenä parempien oppimateriaalien tekemisessä.

Avainsanat: vuorovaikutusmuotoilu, virtuaalitodellisuus, koulutus, oppimateriaali, tutkivaoppiminen

Table of contents

1 Introduction

Finland is one of the leading countries of digitalization in EU level and the citizens have excellent digital skills throughout demographics. Finnish women in particular, excel in the comparison made and rank the first of all groups studied among EU countries. [1] [2] In the *international computer and information literacy* (i.e., multiliteracy) *study* (ICILS) the students of Finland have ranked very good as well and in the latest examination the results of only one EU country were significantly higher [3]. The education sector plays an important role in ensuring people have the necessary digital skills, and the success of Finland is due to the effective and flexible school system [1]. Staying at the forefront of technology requires constant work and like the Red Queen's hypothesis argues you have to run to stay in the same place and to move forward takes double the effort [4]. The hypothesis originates in biology and is associated with interspecies battle. Constant evolution is needed to avoid extinction [5] [6]. Today the hypothesis is being referred to in increasingly new contexts [7] and the Red Queen's hypothesis is very evident regarding digitalization which introduces novel technologies.

1.1 What can VR add to learning?

Providing students with the skillset needed in the 21st century by utilizing simulation and adding elements from gamification can be beneficial [8]. *Virtual reality* (VR) is a new technology and particularly in the field of education it can provide interesting digital learning spaces. Radcliffe et al. (2008) have found that space and how it impacts on learning is besides pedagogy and technology a key element to consider in the learning material design. From design aspect, it is beneficial to utilize the strengths of the technology. The feature that is innovative in VR is its user-centered perspective. The projection is not fixed, but rather the user can move and choose a perspective to view the virtual world [9]. VR enables more student-centered education and with VR, it is possible to visualize something that is not visible in real life, for example, cell structure or atoms. [10] Ethical issues can be avoided by using virtual simulation because no living organism is harmed if virtual 3D models of animals are used, for example, in the laboratory work training [11].

Grimsdale has discovered (1993) that to have a true 3D experience, the interface is required to be a well-established environment that integrates stills, moving 3D and 2D images, 3D sounds, and 3D control. The advantage of using VR is related to its three-dimensional nature and to objects looking similar as in real life. The co-operation is also effective when participants are sharing the same digital space. [12] VR enables digital spaces or virtual worlds where people can interact with each other as digital figures, avatars. Metaverses such as Second life, enable people to spend time, for example, with a common hobby or with educational goals. [13]

The level of enjoyment of the experience can be impacted with design. When the aim is to design IT applications for educational purposes, the usability, the usefulness and the ease of use are highlighted. The usability relates to how easy a system or component is to interact with by utilizing functions available [14] [15]. The paradox of technology is that technology can ease and make life enjoyable, but the difficulty level often increases simultaneously, and the complexity of systems causes frustration for users [16]. A good example is to compare the difficulty of the use of a wristwatch and a smartwatch. Having a standard look in the applications increases the usability and usefulness. A flexible software environment is potentially useful for a wide range of use-cases [12]. Developing a design that is useful requires co-operation from all stakeholders and great project management. [16] From corporation aspect, ideas from participatory design can be utilized [17]. However, it can be difficult to balance against resources and user needs. The product needs to be, for example, usable, understandable, attractive, and reliable to get people to buy and use the product. [16] The aim is to meet the needs of users within limits of money, time, regulations and laws, for example [16].

VR is used in schools from kindergarten to higher education and is seen as promising in terms of learning. It is still a relatively new technology and its degree of adoption especially in primary and secondary schools, has been a slow process [9]. Traditionally, VR has been employed in simulation training, and flight simulators have been among the first VR devices. In the field of education VR has been seen advantageous, for example, in learning languages [18], in sports training [19], in mathematics [20], and in culture heritage education [21]. Furthermore, in medicine VR has been used in the training of surgeons [22]. In culture heritage education, for example, the benefits of VR are increased accessibility, cultural understanding and the possibility of interacting on site [21]. The opportunity for students to develop and design digital technologies and environments is something to consider with digital learning [23]. VR enables more students to participate in learning [11] and it is found that students tend to repeat learning tasks more often when they use VR systems, especially with immersive systems. Whether this is merely students' choice or evidence of inadequacies of the usability design and difficulties to learn can be considered. However, repeating tasks can have a positive effect on learning. [24]

The gamification is something that is discussed today in the field of education. Games have features such as rules to follow, high interactivity, entertainment, usually a competitive element, and a rewarding system [25] [26]. The competitive element can be, for example, points, levels, or time pressure. Games are effective and the effects have been seen in better understanding of content, and at the level of motivation [27]. Repetitive handling of information has been discovered to be a good method resulting in better learning with games. In addition, games are good at building group-work. Games can help to build these skills, even without groupwork designed in, since the shortcomings of usability or the need for solving technical problems, result in people sharing information with each other's at times. [20]

Traditional computer games are designed for only one purpose which is to be played, and completing one task can take a very long time. The user's performance is not a metric in games, and neither is consistency since predictable games become boring quite quickly [28]. Another common feature computer games have is their dynamic and fast-paced nature. [29] As long as the player is entertained, the game is fulfilling its purpose [28]. Possibly because of the nature of amusement of games it has been discovered that some students struggle to take learning games seriously using offbeat nicknames and having somewhat disruptive attitudes. [20] Considering the common features of games, it must, however, be emphasized that the idea of adding gamification and simulation into education is not to change learning materials into games but to add some carefully considered gamifying elements into material to increase students' interest and motivation. The term serious games has been developed particularly for this purpose and can be utilized in case the focus of a game is not to purely entertain but rather to intensify training, for example, in education, healthcare, public policies or communication. [27] [25]

Games have educational benefits and they enhance social, civil, cultural, and artistic skills and are useful in learning information processing, self-reliance and initiative and practicing real world interactions. [20] VR is a platform that games can be developed on. Villena-Taranilla et al. (2022) have found that VR has a positive effect on learning especially with immersive VR, and the effect is similar from kindergarten to fifth grade. According to their findings sessions that last less than two hours are more effective [24]. Many VR solutions exist for school use.

In the VR caves or immersive classrooms, the image is projected on at least one of the walls or the floor of the room (i.e., interactive walls, and floors) [30]. Many schools have obtained VR devices where a smartphone acts as a display and data source of the device. In the desktop VR solution, the content can be viewed with PC. However, the experience on a computer screen is not very immersive but easily accessible. This study is focused on examining the usage of a standalone immersive VR headset that needs no external devices.

In addition to devices, applications are also required and finding suitable VR learning material is important. Ready-made material, virtual meeting places and tools which students or the teacher can utilize to make content for VR learning exist. Applications can be bought from VR stores, for example, the Meta Quest store, and usually, every headset manufacturer has a store. VR applications can be used only with predetermined devices, for example, with Meta Quest-headsets. [31] Meeting applications such as Meta Horizon Workrooms exist enabling a live VR meeting [32] and Microsoft has developed a VR plugin for Teams meetings too to make the meeting experience more immersive. Metaverses such as Horizon Worlds are good for meeting purpose as well [33]. The Second Life was among the first metaverses that have been utilized in the field of education [34]. A corporative EduFinland island existed until 2014 where Finnish schools had a possibility to interact and own a piece of land [35].

For making VR material, exist multiple tools and platforms. For example, Meta Horizon Worlds [36], Engage VR [37], EON XR [38], Second Life [39], VRChat [40], ZappWorks studio [41], Ctrl.studio [42], Wonda [43], Universe (ViewSonic) [44], Eduverse [45] and Roblox metaverse [46]. Educational solutions such as Classcraft and Minecraft Education edition that utilizes gamification in learning are used in schools [47] [48]. The popularity of *Augmented reality* (AR) and VR applications that use geolocation have increased at the beginning of the 2020s in schools [49].

1.2 Challenges of VR

Implementing new technologies into schools needs work, and actions, and it has been discovered that the technology does not work as a catalyst alone but help and backing is needed [50]. The atmosphere and working culture are important, enabling the spread of new technologies and innovations in schools [51] [52]. To be adopted, technology must also fit in with existing practices. Brown et al. (2020) have named the biggest challenge with XR being the need for time and new skills. As a technical aspect, for example, devices chosen and network capacity are very important, impacting the experience of learners. The required immersion can be reached with Wi-Fi 6 and 5G which still has gaps in the coverage and is not available in all places. It is important to notice that the costs of the usage should be similar to those already approved learning solutions to be adopted for action. Brown et al. (2020) argued that with VR the overall costs of usage can be smaller if mobility, fiscal costs and booking and scheduling an expert is more effortless compared with traditional education. Understanding the pedagogical opportunities of this technology is key [50] and debate is whether it is possible to enhance both learning and equity in schools with VR. [11] The technology itself is not enough to motivate students since the novelty fades and thus the content needs to be interesting [53]. Kaisto et al. have observed that assignments that need information search, evaluation and work within a group are perceived challenging but also inspiring among students. The help and *information and communication technology* (ICT) support received has been found to be a determinant in the learning experience which influences, for instance, the motivation and the belief of student's own abilities and is important to consider especially with novel technologies. [52]

Alalwan et al. (2020) have discovered that teachers struggle to find useful material for VR learning [54]. One problem is that material is often designed to be used with a specific device which decreases options available, but also the quality of applications varies [55]. Educational games usually focus on the exercise and practice level and do not offer or require deeper understanding about the subject [20]. In many cases, virtual resources are prioritized and educational content is lacking with VR. For example, in the study focused on the culture heritage education programs, it was observed that the quality of the educational design of assignments was not at the required level. [21]

VR is not broadly used in schools, and neither have many studies been conducted regarding the school usage of VR, but the number has increased significantly lately. If the quantity of research made is compared, VR has not been adopted and studied at similar levels as AR has. [11] A pilot study of the learning impacts of new elements before implementing into the material is a recommended practice. [56] One of the reasons for the lesser amount of research made with VR is possibly that the headsets suitable for the effortless usage in the school environment in technical and financial perspective have not existed long. On the other hand, AR is possible to experience with quite inexpensive mobile devices such as, tablets and smartphones even though more expensive AR glasses also exist. AR technology enables, for example, a textbook to be enhanced to provide audio to listen to, animations to see and queries to answer. [11]

It is notable that VR is utilized with older students more frequently [24]. It appears that in higher education people are willing to take bigger risks and invest more money to experiment with novel technologies such as the potential of VR technologies. As a result of these experiments, for example, Experience Catalogues have been published where VR applications can be searched for educational purposes which benefits the whole education field. Penn States's Experience Catalogue separates different devices and introduces applications for each like Language Lab, Pollinator Park, and Titans of Space [57]. Other projects referred in the bibliography relate to, for example, museums and health care. [11]

In the Chapter 3 is introduced the rationale of the study. This case study focuses on discovering how teachers and students understand VR learning; the important features to implement the material and on the other hand the features to avoid. It is studied the conditions that enable VR learning in schools and practices to design effective and useful learning materials. Chapter 2 on background, introduces the three main areas of the study: the education, VR, and design in more detail to have a strong theory base and to understand the context. Data is collected through questionnaires as described in the Chapter 4 and from user tests presented in the Chapter 6. The Chapter 5 introduces the VR application specifically developed for this purpose and utilized in user testing. The analytical focus of the study is to find whether there exist a theme or qualities that are seen or experienced more important than others to be included in VR learning. Furthermore, can there be found connections between responses and demographics (i.e., the age) and is there a difference how students and teachers see VR learning. The aim, as found in Chapter 7 and Chapter 8, is to test a design model suitable for the purpose and to reflect on what teachers and students consider the standards of quality in VR learning materials.

2 Background

The aim of research is to deliver better designs and *information technology* (IT) artifacts [16]. Skepticism and empiricism based on observations are in the focus as well as data, publicity, adversarial and peer reviews. Instead of relying on intuition, authority, illusory correlations or pseudoscience, the foundation of scientific research is the results of the experiments. [58] This theoretical background part begins with a short introduction to the research framework of this study. Additionally, the two main trends of design are discussed.

Research of IT solutions can be performed from different viewpoints: *market research* or *design research* (DR). In market research, the interest is what people buy and how they make their purchase decisions. On the other hand, the focus of DR is to identify what people's needs are and how they use products. The data received from marketing research (quantitative research and "Big data" e.g., marketing analytics) however, do not reveal people's actual needs, desires or the reasons behind their decisions and behavior. In DR, qualitative observation methods are normally used, and the sample size studied is rather small, dozens of people because of the time consumption. The research focuses on people, their decision-making process, and related environmental factors. However, to make sellable products, both marketing and design research are needed since it is important to know what people buy as well as what they actually need. [16]

Originally methods of DR focused on building *information systems* (IS) and excluded evaluation, which is currently recognized as an essential part of the design process revealing inadequacies related to interactions. *Action design research* (ADR) is designed to fulfil this by adding a layer where the interests of organizational context, for example, developers, investors and users are included. The ADR methods address the problem, construct, and evaluate the artifact, reflect and apply the solution to similar use cases, and finally formalize the outcomes as design principles. [59]

This study is utilizing *Design science research* (DSR) practices. DSR differs from earlier research by adding a layer of scientific knowledge on IS design. It focuses on developing IS and the processes supporting emerging knowledge. [60] [61] Seven guidelines to follow in DSR are: (1) produce a viable artifact, (2) solve a real life problem, (3) evaluate (e.g., field study), (4) make clear research contributions, (5) follow methods rigorously, (6) find available resources and knowledge (e.g., frameworks, models and methodologies) and finally (7) present research finding to the public. [62] [63]

The following Chapter 2 describe the theoretical background (i.e., knowledge base) of this study. The study is a combination of three key elements: education (2.1), VR (2.2), and design (2.3). Education and VR are environmental components, setting the boundaries for the design. Education is considered an organizational factor and VR as technological. Each component provides a slightly different point of view for the study introducing elements currently relevant for the theme. The education theme, for example, introduces suitable learning approaches and challenges of digital learning and learning material design. The virtual reality theme introduces the qualities and possibilities of VR, concepts important to recognize, and devices used. It offers insights into what VR is and can be used for in the school environment. The last theme that is focused on in the study is design. The special interest is humans and *interaction design* (IxD) and how good learning materials can be designed.

2.1 Education and digitalization

We will start by introducing the education component and discover how computers first arrived at schools since VR headsets are also wearable computers. The special interest is how teachers and students see digitalization. Attitudes, interests, motivation and immersion components are presented. After that, it is discovered what are learning materials, e-learning, pedagogical quality, learning approaches and learning spaces.

2.1.1 From the arrival of computers to 21st century skills

Computer science is for the first time mentioned in the Finnish national curricula in 1985 and computers were adopted into Finnish schools mainly in the late 1980s [64] [65]. At that time computers were becoming popular among ordinary people [66]. The digitalization process has been gradual and appropriate in the field of education [67] and speed of implementation and opportunities to use digital devices has varied and varies between schools even today [68]. According to Kaarakainen and Kaarakainen (2018) three phases of educational digitalization exist in Finland: (1) acquisition of devices (1998 to 2004), (2) evaluation of benefits (2005 to 2010) and (3) diversification of use (2010 to 2018) [69].

Education has guidelines to follow. Education is regulated with legal framework and learning objectives are based on common education goals aiming to bring up active and independent citizens who respect the values of society, for example, social, ecological and economic sustainability, confidence, fairness, responsibility and transparency [70] [71]. Today digital participation, equality, and civilization are defined as specific goals related to digitalization [71]. Technical devices and new educational materials have been adopted into education throughout the history [72] and after computers were invented, there has been an increasing need to integrate digital technology into education. Separating technology and pedagogy can be difficult and even unnecessary today [73] however, pedagogy is something that has the highest impact and explains learning more than the technology used [68].

Digitalization is a common challenge as employment and daily life in general require digital skills [74]. The meaning of the term digitalization is, however, not yet established [75]. The digitalization of education can be identified not only as an implementation process of digital devices, but also as pedagogical changes needed to enable the implementation [76]. Critical thinking and the ability to evaluate the reliability of information are valuable skills in the 21st century. Finding reliable information can be a challenge since search algorithms provide a modified truth. [77] Multiliteracy, which is a skill needed today, includes not only reading regular textbooks but the knowledge and use of different digital devices (e.g., VR headsets), applications and services [78].

Finding suitable applications and digital services is a challenge for schools and teachers and adopting, for example, a new learning management system can be difficult, and needs lots of work [79]. Another issue raised during the COVID-19 and needs attention is how digitalization has turned basic education into a marketplace which may be overwhelming [80]. Data security and data protection are very important concepts in digitalization today as well and need to be considered well in schools [81].

Teachers and digitalization

There have been many challenges faced during the adaptation process of computers into schools, for example the lack of teacher training in quantity and quality and the lack of computers. [66] To provide all students equal opportunity to learn digital skills, teachers' mastery of digital skills is important [82], but the lack of scientific consensus and not having unambiguous proof of the benefits of digital devices as well as the lack of common goals have been confusing [83] [78]. In the near history, digitally enhanced learning may have been seen as an alternative learning method instead of as a compulsory element to include. Teachers have struggled to find meaningful pedagogical use for digital technologies in history even though the large majority have had adequate technical skills for the usage. According to the latest ICILS report three out of five teachers in Finland considered not having enough time to prepare lectures utilizing ICT [78]. Having compact ICT projects, receiving help from external experts and collaborative teachers has proved to extend ICT skills of teachers. [84] Furthermore, ready-made learning materials have been seen beneficial [85].

In some countries computing has a status as an independent subject, however in Finland the goals are embedded inside other subjects which can be beneficial and emphasizes the multilevel use of technology. Computing is often offered students as an optional subject in lower and higher secondary schools [78]. In the Finnish national curriculum (2014) ICT skills are included as a part of a comprehensive set of skills. However, because of that the responsibility of teaching ICT is not clear, and for example, evaluation has appeared to be challenging [86] [87]. Descriptions of digital competences published in 2024 are intended to demonstrate the student's destined learning path and skills needed to ease the confusion [88].

Students and digitalization

The discussion about the competence and skills required of students in the 21st century has been intensive. The issue every framework agrees on is the importance of the knowledge of ICT [74]. The technical resources of both home and school environment have been very good from the beginning of the 21st century. Students are motivated and capable of adopting new technologies. [84] However, learning outcomes with ICT are controversial since the self-regulating skills of students have significant impact on learning and the role of these skills are in many cases highlighted when ICT is involved [89]. The current conception of learning emphasizes the active role of a student and the collaboration aspect in learning [86]. Implementing ICT as a part of educational practices enables more student-centered, collaborative, and inquiry-oriented teaching, emphasizing students' responsibility as a learner [84]. However, students are at the same time concerned that reshaping learning decreases teaching which they see valuable [53]. In addition, using digital devices, teachers have little control, for example, where students search for information, and at the same time what connections to the outer world they have [90]. Not only the devices acquired are important but how devices are utilized as well. Digital multitasking (e.g., messaging during lecture) is something that disturbs concentration. However, the use of digital materials that are well-designed has a positive effect on learning [91] [68] [92]. The challenges of VR can be very similar to any use of a computer in school.

The Finnish national educational policy strongly encourages implementing ICT into pedagogy. One of the goals of technology usage in schools is to promote parity and equality among students. [93] It is important to ensure that students know how to use new devices and how to act in digital environments [94]. Linguistic, learning and motivational difficulties that require slowing down progress in education path are factors that are proved to endanger the development of digital competence and expose students to be excluded from digital society and to be more vulnerable, for example, to fake digital information [95]. Accessibility is also something to consider with equality, and people with specific linguistic difficulties suffer from inadequately designed digital services struggling to find information. Clear and consistent structure and texts, paragraph divisions, simple font, relevance and easy to read texts, as well as safety and reliability are important with accessibility. [96] Performing pilot studies and user tests for services before delivery is important [16] [56] [97].

2.1.2 Learning materials, e-learning and pedagogical quality

It is not easy to define the concept of learning materials since the diversity of materials is great. According to Barker and Campbell (2010) learning materials are *"anything used for teaching and learning"*. Material can become learning material after it is approved by teachers and used in learning. [98] E-learning materials are, for example, simple digital assignments, simulations, illustrations, thematic entities, assignment containers, a course or a part of a course, extra material, or content meant for the teacher. Utilizing social media facilities and different software, for example, Microsoft Word are important for developing students' digital skills as well. [54] Learning materials can be divided into three types: (1) *functional learning materials* (e.g., tools, devices, computer programs, LMS), (2) *semantic learning materials* (e.g., texts, images, videos), and (3) *"didacticized" learning materials* (e.g., textbooks, educational games). Didacticized learning materials have inscribed planning and expectations of what activities the students and the teacher are aimed to perform. [99]

ICT has the potential to enhance learning when the quality and design of material is good. [100] [101] The suitableness of material for teaching and learning purposes, making teaching and learning easier and producing extra pedagogical value are in the center of pedagogical quality. The extra pedagogical value can be diverse opportunities delivering or making assignments, the creative use of knowledge, and the novel sense of community. Utilizing the latest research findings and supporting teachers to embed new technologies (e.g., VR) is important. Adaptable materials are most usable since enabling versatile use serves the heterogeneous group of teachers and students better. The high-quality learning material supports learners' conscious thinking processes and enhances active participation. [54] [102]

The pedagogical purpose has two main aspects. It refers not only to the learning content but also to the learning process required of the student. [70] Traditionally pedagogical interaction is seen between three components student, teacher and material. The pedagogical purpose of teaching is to facilitate the interaction between learning material and the learning process of the student [103]. Teaching is not random but is based on careful planning framed by curriculum, teaching methods, and the class schedule. Learning can happen spontaneously and subconsciously, but studying is always an active process. [70]

Attitudes, interests, motivation and immersion

It is important to consider the attitudes and motivation of learners when designing educational materials. Once students are interested, the learning experience can be more enjoyable, which can impact learning outcomes. [104] Interest and *intrinsic motivation* are significant affecting the engagement of the learner and resulting in positive behavior [105]. Interest relates to the emotions and values of a person [106]. In intrinsic motivation, the interest can be situational or individual interest. *Situational interest* can be triggered without earlier knowledge of the subject since curiosity and enjoyment are determinants [107] [108]. *Individual interest* refers to a situation where a learner has a deepening relation to the subject, and the task itself has a minor significance for the engagement. Personal values and continued engagement are at the center of this concept. [109] [105]

Situational interest is something that can occur regardless of students' individual interests [107]. However, students experiencing individual interest are often more focused, engaged, and receive better grades [110] [111] [105]. Maintaining interest is a challenge while interest can withstand varying periods of time [112]. Immersion is a state of mind where someone is feeling surrounded and being part of something physically or virtually, immersed in the experience. There are three types of immersion recognized: imaginative (the willing suspension of disbelief [113]), challenge-based (i.e., the flow [114]) and sensory immersion (the cinema of attractions [115] e.g., audiovisuals). [116] When immersion is more traditional style it is called the willing suspension of disbelief and the experience is more passive, for example, viewing a movie at the cinema [113]. With flow, more active engagement and adaptive processes are expected. Flow is balancing with the level of challenges and the skills of a person involved [114]. In sensory immersion, audiovisuals are in the center. Multisensory

VR environment can be seen as an example of this kind of immersion [116]. On the other hand, VR experience can be a combination of all of these three types of immersion.

The individual interest in digitalization and learning varies between students, and it is important to consider in learning material design. Koivuhovi and Polso (2022) found in a study conducted in Finnish lower secondary schools that students can be classified into four profiles based on their attitudes on learning and digitalization as illustrated in Figure 1. Two thirds of students have high digital competence and positive attitudes towards learning and digitalization ("Know and interested"). A sixth of students' attitudes are more moderate and neutral ("Neutral interest"). In the third group ("Know, but not interested") a seventh of students have high digital competence, however their attitudes on digitalization are more negative. The students in the final group ("Don't know nor interested", 1 %) have the weakest digital competence and negative attitudes on digitalization. [117]

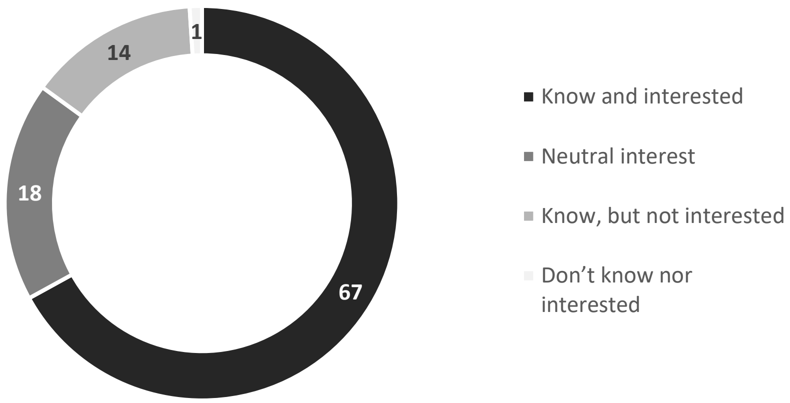


Figure 1. Digital profiles (%) of Finnish lower secondary school students (n= 6936) by Koivuhovi and Polso [117].

In the same research project Hotukainen and Oinas (2022) found a significant difference regarding student's age and enthusiasm related to digital devices. It was discovered that the enthusiasm towards digital devices decreases every year in lower secondary school and the concern is that negative attitudes on digitalization possibly have an impact on overall learning as well. Some students also suffer from anxiety towards digital devices. A fourth of the students suffer from anxiety at some level and every seventh of them experience obvious anxiety when using digital devices. [118] These findings are something that are important to consider when designing digital learning experiences and materials for students. The material should be beneficial for every student.

2.1.3 Learning approaches and learning spaces

Many approaches can be utilized in the design of educational materials. In the following, are introduced a few currently influential approaches to learning. The introduced types are *inquiry learning*, *the invention learning approach*, *problem-based learning* (PBL), *project-based learning* (ProjectBL), *phenomenon-based learning* (PhenoBL)*,* and *trialogical learning approach* (TLA).

In inquiry learning the goal is to get students to think like scientists. They get tasks that aim to advance their knowledge in a particular area. For example, how to measure friction between a floor and a shoe. The approach has predetermined goals that students are aiming at. [119] [120] The idea of the invention learning approach is that students are solving a challenge by inventing a solution. This approach emphasizes creativity, has an object-driven nature, engages student's curiosity, teaches collaboration, has a nonlinear nature, and is suitable for all ages. Maker space is also a term connected with invention learning. [121] [122]

PBL has been seen especially advantageous in the fields of medicine and computer science [123], but the lack of individual learning support has been seen as a problem [124]. In the center of this approach is the activation of the prior knowledge, connecting learning into situations that are feasible in further career and fostering elaboration of material [125] [123]. In PhenoBL observations are in the center. Students are asking questions, for example, why airplanes stay in the air and do not fall. They start to determine the reasons behind this by combining their current knowledge with new information. PhenoBL is connected to real life observations and problems enabling critical thinking, problem-solving, communication, and collaboration skills to develop. PhenoBL can also be considered a subclass of problem-based learning. [126] [127] [128] VR is found a valuable tool to implement ProjectBL and increases motivation and problem-solving skills. The advantages of ProjectBL are collaborative learning and co-construction of knowledge. [49]

TLA is a relatively new approach. It is designed to be used with older students, and the collaboration with working life representatives adds value to this approach. There are six design principles to include, for example, "*to emphasize development and creativity through* *knowledge transformations and reflection", "to promote cross-fertilization"* and *"to provide flexible tools for developing artifacts and practices"*. [129] [130] [131]

All of these mentioned learning approaches have certain benefits. This study aims to utilize the inquiry learning approach in a learning material design process. The approach is especially beneficial regarding science learning that is the focus of this study.

Learning spaces

In the design of educational material, not only is choosing a learning approach important but the space where learning is situated as well. The learning space is, for example, a classroom but can be any space where learning takes place. Designing *Next Generation Learning Spaces* (NGLS) [132] that enable, and ease learning is a complex process that requires intensive collaboration and participation of all stakeholders. Innovation and experimentation are in focus with NGLS as well as changes in social patterns and funding, moving to more learner-centered pedagogy and implementing novel technologies such as VR. Designing a learning space can be seen as a dialog between pedagogy and technology and aims to meet different needs. For example, the JISC (*Joint Information Systems Committee*) report focuses on qualities such as flexibility, future-proof, boldness and supportiveness [133] [134].

Slotta (2010) considers future classrooms as an interplay between pedagogy, technology, and community. "The Smart Classroom" research environment needs new approaches on learning and instruction that enable innovative experimentation climate, with collaboration and exchange between researchers. The wide community benefits results. The *Web-based Inquiry Science Environment* (WISE) on the other hand emphasizes the educational innovation aspect. [135] The idea of the *biology guided Inquiry Environment* (BGuILE) is to support students' scientific practices and enable inquiry learning culture in a classroom with technological aids [136].

To facilitate the design process and evaluation of learning spaces, Radcliffe (2006) has generated the *Pedagogy-Space-Technology (PST)* framework based on the findings of the NGLS project. The three key components: pedagogy, space and technology cannot be put in a hierarchical order but should be seen as having continuous reciprocal dialog and shaping each other. However, pedagogy can be seen as an entrance point to the framework. The key elements of PST framework are collaboration in all phases and continuous reflection on what each stakeholder is processing and for what purpose. To deliver valuable learning space, each participant has a certain important role and a reason to contribute to the process. [134]

Traditionally, the learning space introduced in PST framework has been a classroom. However, novel technologies enable new opportunities. Learning space can today be defined to be virtual spaces as well, and to offer immersive experience that can be accessed with, for example, a VR headset from different physical places. This study utilizes the ideas of PST framework. The next section focuses closer on VR technology and the possibilities it has for learning.

2.2 Virtual reality

Next, we move on the second component VR, and begin the journey from ancient Greece, as other realities have from early times fascinated people, and human perceptions are at the center of the dilemma. Plato’s Allegory of the Cave was the first consideration of parallel realities. People living in the cave and seeing the world only as reflections and shades on the wall are unaware of the world outside the cave. Their only reality is the cave and entering the world outside seems a risk not worth taking since requiring the alteration of personal perceptions. [137] Later philosophers have had considerations of human perceptions and what is real. *“Cogito, ergo sum — I think, therefore I am”* is in the center of René Descartes’ philosophy. He doubted the signals received from our senses. However, a person’s existence is something that is never doubted. [138] Is anything real or are our senses fooling us? This is the question to ask.

The idea of VR is that we are sensing and experiencing something that is not real but digitally developed. When it comes to VR, devices are often the first thing that comes to people’s mind. In the early 2000s people associated the concept of VR strongly with devices that were used in VR experience: head-mounted devices (e.g., goggles) and sensing gloves [139]. The field of the technology is fast-changing, and it is not beneficial to rely on qualities of a specific device to define a system. The definition needs to be universal.

VR can be considered to have its origin in the era of early cave paintings when spectator’s imagination played an important role, because of the lack of details in paintings. The details of images gradually increased throughout history and adding movement to images in the forms of animation was an important milestone enabling animated films and finally television to develop. [13] Seeking ever more immersive experience finally led to the development of

VR systems. The history of VR devices is not long but it is fascinating. Devices such as view-master existed in the late 1930s with still images [13] [140] but the development of VR devices can be considered to really have started in the USA in the 1960s. The process was gradual, and the first VR conference was held in France, in 1992 and IEEE *(Institute of Electrical and Electronics Engineers)* arranged another a year later in the United States. After that, the field of VR can be considered to be properly acknowledged. [139]

2.2.1 Virtual reality definition and related concepts

The first person known to use the term *virtual reality* (VR) was philosopher Immanuel Kant in the late 17th century, but he was not referring to it in the way that is familiar to us today but using it to refer to a reality inside someone's head opposite to physical world [141]. The modern concept of VR was developed among NASA's scientists around 1985 after a name for a new system which they were developing was needed. However, it took almost a decade for the term to get established. The *Virtual Environment* (VE) was a rival and was used as a parallel concept, especially among academics since the term VR was considered vague. A rival concept was *Artificial Reality* used in Japan during the same era. After the wide audience preferred and was more familiar with the term VR, it became generally accepted round 1992. [142]

Similarly, there has been a debate about what the definition of VR should include. At the time, it was even used as a synonym for artificial intelligence. Therefore, finding a comprehensive definition for VR became vital. [142] Multiple definitions for VR have been made in history, and each one has had a slightly different focus. In the following, five definitions from the early 1990s to the 2020s will be introduced.

The first definition is Burdea's (1993) Virtual Reality triangle (I^3) where components: interaction, immersion, and imagination are present. The user interacts with the content and is immersed in virtual surroundings. The term "imagination" refers to the imagination and skills of the designer, which the content the user is enjoying is dependent on. Four key features of VR are, the three-dimensional world, computer generated content, interactivity, and physical presence simulation. [139] [143] Benford et al. (1996) have defined Virtual reality through dimensions of transportation, artificiality and spatiality. In immersive Virtual reality, user is fully involved, and isolated from the physical environment, and immersed in a computer-generated reality which is also a collaborative spatial system. [144]

Bryson's (1998) definition states: *"Virtual Reality is the use of computer technology to create the effect of an interactive three-dimensional world in which the objects have a sense of spatial presence"*. The aim of this definition is to separate the concept from telepresence (e.g., person being present real-time through a display of a tablet connected to a robot) and animation, define the experience to be more than an illusion, and underline the key element, the spatial presence of an object. The spatial presence refers to the object being fixed in a virtual place in the digitally generated environment similarly as, for example, a chair is in a real room. The image on the screen changes gradually when the user is turning their head but the chair stays in the same place in the room (compare head-mounted television). This feature is gained through the calculations and counter rotation of the image and needs real-time user tracking. According to this definition, immersion is not needed for VR experience, and the environment does not need to resemble the real world. Neither is it important to have visual fidelity (i.e., sharp image). However, functional VR systems need to have a close to real-time performance and a sufficient user tracking system. [142]

In the fourth definition, new aspect is to specify that multiple sensorial channels are affected during the experience and emphasize the role of simulation and the interface aspects. According to Burdea et al. (2003): *"Virtual Reality is a high-end user-computer interface that involves real-time simulation and interaction through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, smell, and taste"*. [139]

The latest fifth definition introduced here is from LaValle (2020). VR is: *"Introducing targeted behavior in an organism by using artificial sensory simulation while the organism has little or no awareness of the interference"*. LaValle's definition aims to be as universal as possible in the same way as the earlier ones. The use of VR has changed over time, and an update for the definition is needed. LaValle has chosen to use the word "organism" since VR experience can be designed not only for humans but for other organisms as well, for example, for gerbils to study neuroscience. The targeted behavior refers to something that the designer has intended the organism to experience, for example, to fly. The "fooling senses" is present in this definition claiming that there is little awareness of artificial interference. [13]

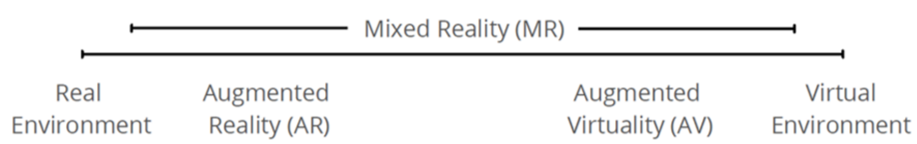


Figure 2. Milgram and Kishino's Reality-Virtuality (RV) continuum from 1994 [145].

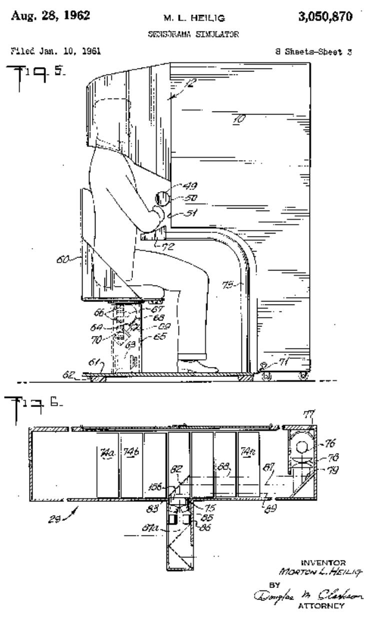
Today there is debate whether the term VR is useful anymore because the industry has gone through profound changes. In 1992, a new term *Augmented reality* (AR) rose [146] and today there are many terms used referring to virtual systems. The term "VR" is usually referred to in the context of fully synthetic and systems isolated from the real world. Everything that the user sees is artificial in the environment, and an illustration of an alternate world. [147] [13] In AR the user is experiencing the real world with place-bound digital content. The digital content can be, for example, texts for guidance or artificial objects enhancing human perceptions. Third term used today: *Mixed reality* (MR) is very controversial and many people consider MR as a synonym for AR. Some are defining it to differ from AR based on interactivity. In MR application, users can manipulate and walk into the scene. With traditional AR, it is only possible to view static content. [147] Mixed reality can also be seen as a continuum between VR and the real world as can see in Figure 2. It is a mixture of the real and the virtual environment. [145] The use of the terms VR, AR, and MR, has decreased recently because of the introduction of novel technology that enables the use of multiple technologies (VR/AR/MR). The term favored by professionals today is *Extended Reality* (XR) or X Reality which covers VR, MR, and AR. However, a combination of terms VR, AR, and MR, are also used such as AR/VR. [13]

When we talk about VR systems, it is important to mention *metaverses*. According Wang et al. (2023), the term doesn't refer to technology, but an environment that uses XR technology. Metaverse is an upgraded 3D version of the internet. An immersive and spatiotemporal virtual place for people to meet, socialize, work and play as avatars (e.g., digital selves). The terms: second life, 3D virtual worlds, and life logging are also used. [148] Immonen et al. (2025) have identified six elements that are mandatory for metaverse. Metaverse is online and synchronous virtual space, which enables real time interaction, collaboration and presence as avatars. [149]

Concepts of digital VR systems are in a constant transient state, and the industry is in change, novel technologies are adopted, and the popularity of systems is steadily increasing. In this study, VR is defined by the latest definitions from LaValle. The term VR is used in the study instead of XR since the aim is to study qualities common for VR systems and devices especially.

2.2.2 Prototypes and later models

The idea for VR devices originates from Morton Heilig who was a cinematographer. He invented and built in the 1960s the first system that can be considered VR. The system is called "Sensorama" and it is illustrated in Figure 3. Sensorama is a head-mounted television with 3D, wide vision, motion, color, stereo-sound, 10 aromas, wind and vibrations [150]. The user sits on a chair, head pressed inside a multisensory box, viewing a movie in 3D, smelling aromas, hearing sounds, feeling the wind created by fans and sensing the bumps of the road through a vibrating chair. Spectators can choose the experience out of six films. For example, riding a motorcycle through New York. Morton's project was pioneer work and however, because of the lack of funding it ended up unfortunately quite soon. Heilig also made another invention related to VR which he patented as well. His drawings of a Telesphere mask (a head-mounted television), also illustrated in Figure 3, resamples very much VR headsets used today. This mask had 3D effects, controls, and optics to focus on the image, stereo sound, and smell capability. This invention did not reach the attention of the wide audience either during Heilig's lifetime. [151] [152] [139]



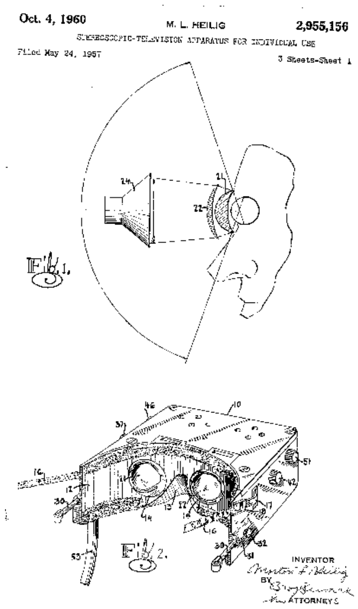


Figure 3. Morton Heilig's VR inventions. On the left side is the Sensorama from 1962, which is considered to be the first VR device. On the right side is the Telesphere mask from 1960. [153] [154]

Ivan Sutherland worked with early VR devices as well and can be considered to have continued Heilig's work. He built a *head-mounted device* (HMD) composed of two miniature cathode ray tubes (CRTs). In the device, the user can simultaneously see digital images and the actual objects of the room. The device is connected to a mechanical arm which tracks user movements. Since the first mechanical tracking system was very heavy, Sutherland and his colleges built an improved model with a lighter tracking system based on ultrasound shown on the right side of Figure 4. The system had three transmitters added to the HMD. On the ceiling there were four ultrasound receivers which tracked movements and sent the data to the computer which updated scenes displayed on the HMD. [155]

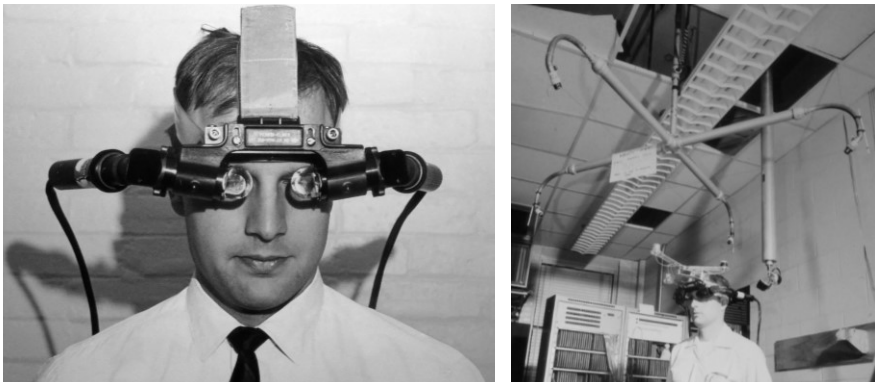


Figure 4. Ivan Sutherland's head-mounted display from 1968 [155].

Sutherland's HMD used a scene generator which he and his colleague developed in the early 70s. It was a precursor for modern graphic accelerators where computer generated scenes eventually became animations. It was a novel way of generating images since Heilig's Sensorama was based on analogical video images [139]. The graphics were quite simple, and the device could display max 400 polygons with 1/20 sec frame rate. Sutherland also predicted that haptics and gaze control would be added into VR experience later. [10] [139] Frederck Brooks Jr and his colleague executed Sutherlands idea of haptics a couple of years later [156] [139] and developed a "joystick" that gives force-feedback and mimics, for example, the feeling of flying an actual airplane [10].

The military was especially interested in VR simulators in the 1970s, and the following decade was dedicated to developing VR for military purposes. Diverse simulators were built, for example, VR flight simulators. It was easier and more inexpensive to update a VR simulator's program to match a new airplane model compared to building a whole new analogical simulator. [139]

NASA became interested in VR simulators to train astronauts in a space-like environment in the 1980s. An LCD-based head-mounted device (VIVED, 1981) was made by reassembling existing devices like a portable TV, adding optics, a host computer, a graphics computer, and a tracker. Later, a sensing glove was developed and added to the experience. NASA's scientists also created software for 3D virtual sound systems. VIEW was developed after VIVED and utilized already more powerful computer and had improved graphics. [139]

Commercial VR products

After Heilig's Sensorama and until the late 1980s, VR devices developed were designed for prototyping, private or governmental use. In the late 1980s, a company called VPL started to sell VR products for regular people. Their products were EyePhone (HMD, price of $11,000 (1987)), DataGlove (a sensing glove – substituting a keyboard and a mouse), and Reality Build For Two (multiple users can interact in the VR environment simultaneously). [157] [139] Nintendo released a sensing glove product at that time also [143]. PowerGlove was the first VR product that became a commercial success. However, because of the lack of suitable games, PowerGloves were not produced after 1993. [139] At the beginning of the 1990s devices such as a stereo visual system, gloves, wands (controllers), a 3D mouse and tactile displays were developed [12]. In the 1990s video game arcades were popular and wide audiences were able to experience VR games there (e.g., Virtuality by SEGA) [13]. A commercial VR workstation: Vision was published in 1991 and later came updated models ProVision and SuperVision (prize of $70,000 (1993)). [12] [139]

During the 1990s costs remained high, but the resolution of the image and sharpness slowly improved while introducing, for example, Video Graphics Arrays (VGA) in 1997. For example, Sony introduced Glasstron (HMD, weight 310 g) and Kaiser Electro-Optics implemented Extended Graphics Array (XGA). Characteristic for the 1990s was also large-volume displays and wall-size images through which multiple users could take part in the same VR experience. [139] This technology was called the *CAVE*, and was originally designed to boost scientific discoveries [158] [159].

As we move closer to the present, components that build VR experience have become more established. For example, (1) a visual display (e.g., lighting, textures, reflections, and shadows), (2) head- and hand-position sensing, (3) force feedback (useful in e.g., surgical training), (4) 3D-sound (to add more realism and voice control (input and output)), (5) other sensation (e.g., vibrating and tilting seat) and (6) co-operative and competitive elements compose the VR experience. [160] Devices in the early 2000s used magnetics or ultrasound for tracking. Today's trackers use generally the wavelengths of visual light or infrared [161]

[162]. The value of the VR industry increased from the early 1990s to $50 million [163] and was at the beginning of the 2000s at $1.4 billion [164].

In the 2000s, HMDs were still highly dependent on powerful computers. After the mass spread of smartphones, it was finally possible to design affordable devices for a wide audience. Many advances made the growth of VR possible: advanced game engines, increased rendering power and smartphones that had high-resolution displays and sensors for sufficient tracking. The first versions of these headsets were VR cases that required a smartphone inside to be the display and the computer of the headset; however, lenses are included in the case. Devices are, for example, Google Cardboard [165] and Samsung Gear VR [166].

In 2010s, for example, Microsoft, HTC and Oculus developed VR technology [167] [168]. Oculus was the first company that made VR headsets that utilized embedded android technology resulting into lightweight and affordable headsets. The VR headset was based on Palmer's prototype from 2012 [169]. The first product built for commercial use was Oculus Rift and was followed with a set of models, for example, Oculus Quest 2 and 3. [13] The Oculus's system uses android's affordable MEMS sensing and video display technology. The result is a rather wide field of view and a more extensive and immersive experience. The technology was originally designed for the game industry's needs but has been adopted in various sectors since. Oculus's headset enables not only gaming but also telepresence, travelling, and virtual attendance. [169] Headsets are useful also for educational purposes. In 2014 Facebook (later Meta), bought the Oculus VR company and started to develop virtual reality environments [170]. Many companies followed, and in a short period of time, headsets that were using this technology were produced widely. For example, Sony developed a VR headset for PlayStation in 2014 [171].

Other competing technologies existed simultaneously. For example, HTC released a VR headset based on PC technology in 2016 [172]. The outcome of this competing set-up was, for example, to get more powerful devices but also to lower costs after companies were forced to search for more affordable technologies. However, how to utilize VR, and where to use the VR technology for, was still not clear [173]. Today devices are used in different fields and in the healthcare industry, for example, it is possible to treat disorders with VR such as anxiety. The method is called *virtual reality exposure therapy* (VRET) and is found to have better treatment gains in the long term compared to traditional exposure therapy. [174]

Waiting for the breakthrough

From the early beginning people's experiences of VR devices have been impressive even though technology has not been very efficient [142]. VR has been mentioned to be "the next generation of *Human Computer Interface* "(HCI) [12]. VR has *"the stigma of unkept promises"* [13]. The potential of VR systems has been recognized for a long time, but the adaptation process has been slow for several reasons. The biggest being that technology has not been advanced and powerful enough and has required lots of space [139].

The development of the field of VR is highly connected and dependent on the development of the technology [12]. Adequate performance of a technical system is needed because it is important for the experience for changes to happen in real time. In the VR systems, visual fidelity and sharp resolution have not been the determining element in the VR experience. For example, in 1988 the sharpness of NASA's VIEW was poor as well as Sutherlands' earlier version 1967, but even still people were amazed. [142] However, when technology develops, people are expecting more from devices.

It has been discovered that to gain people's acceptance for an invention development needs to be more evolution-style rather than a revolution. In 1990, it was realized that not only devices need to improve but also software. Computer systems need to be capable of processing VR applications, and the software needs to be flexible. Also, suitable VR development essentials are needed to make the content and applications. It is important for the software to be compatible with existing systems or have extensions to connect software to VR. Creating everything from scratch is time-consuming and puts a lot of pressure on software developers. [12]

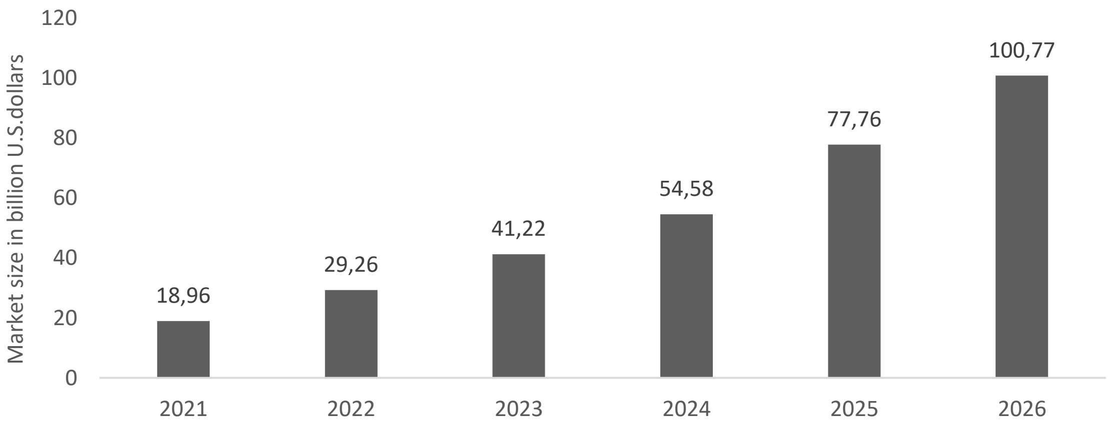


Figure 5. The Worldwide market size of extended reality (XR) [175].

The development and research phase has been taking a long time, but all the building blocks for VR breakthrough were considered ready in the 1990s and there were grand expectations towards VR. However, the technology did not reach mainstream and the expectations of people [13]. The field experienced a crash. The VR industry was hibernating until the 2000s. Technology that was in the center of new revolution was Android-based Oculus VR headset. Whether this is the final breakthrough for VR that the world has been waiting for or what will happen in the following decades cannot be known. However, as can be seen from the chart shown in Figure 5, the market size of XR has been steadily increasing, and the growth is expected to continue in the following years [175].

2.2.3 Devices and applications today

Multiple devices to experience VR exist today. Models and especially the features of devices greatly determine the user experience and the degree of interactivity. VR content can be viewed with head-mounted VR/XR headsets, smartphone requiring VR/XR devices, CAVE system, on PC's or on mobile devices screen. Head-mounted devices offer the most immersive experience [176]. Depending on the device and the model, the utilized I/O (i.e., input output components enable interaction between the user and the application) components vary, which is presented in Figure 6. Most commonly, the headset is a visual output device with speakers since sounds are also a significant part of the user's VR experience. Other modalities and senses can also be stimulated. In the earliest models, for example, the sense of smell was also included [139], but with later models, are rarely used. Most commonly hand controllers are user input devices, but hand gestures can also be used in some applications as an input. Haptic gloves can be used with specific devices. It offers an experience of touch when interacting with objects. [177]

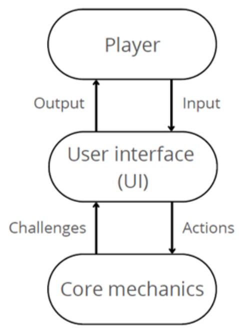


Figure 6. Design components of a game by Smed et al. [178].

Many schools have acquired VR models that require smartphone inside. These models have lower price tags from €20 to €50 and enable viewing VR content, for example, 360 videos. In addition to this, some models have interaction option with a hand controller(s). [179] In the same way, inexpensive XR devices (price €100) have been developed with the option of interaction. These devices need as well a smartphone inside [180].

VR CAVE systems are the most expensive VR systems (€80 000 to €750 000) and are based on images that are projected onto walls. There are glasses as well to wear and a controller for interaction. [181] There are also more inexpensive modifications available of VR Cave systems for educational purposes, for example, immersive classrooms, walls or floor, where images are projected only a few walls or floor or used big screens [182] [183] [184]. Many VR applications (e.g., Meta Horizon Worlds) can be accessed also through regular internet browser and viewed with a PC's screen. This desktop VR is the most accessible and requires no extra equipment, however, the experience is less immersive for that reason. [36]

VR headsets are today more and more standalone devices instead of requiring a cable connection with a powerful computer. VR devices sold are, for example, Sony PS VR, Bigscreen Beyond and MeganeX and the price varies between €480 and €1700 [185] [186] [187]. Many new VR headsets have also ability to view AR content. This is enabled with passthrough cameras (i.e., passthrough AR). These new VR glasses are called XR headsets (e.g., Meta Quest 3 & 3S, Vive XR Elite, Pico 4) and cost from €350 to €900. [188] [189] [190] To choose suitable headset for the purpose is important, and many headsets sold today for customers are designed for playing VR games. In this study is used standalone Oculus/Meta Quest 2 VR headset model.

2.3 Design

The third component introduced is design. The aim of design is to deliver products that meet the requirements and core needs of people. To understand the needs, also the underlying, and developing the best possible product is a task of a designer. To achieve the goals are multiple ways. Problems with a product or user experience, the problem space is solved in the design space. The design process includes many representations also known as models. Models can be mental (i.e., conceptual or theoretical), illustrative such as scenarios or physical prototype models. [191]

To solve the problem space can be used practices of *design thinking.* Design thinking is a process to find and solve fundamental problems (e.g., core needs) and can be utilized not only, solving design related but other problems as well. Based on Norman (2013) *"A brilliant solution to the wrong problem can be worse than no solution at all: solve the correct problem"*. Finding the real problem and solving that is what a good designer does. This is accomplished by studying people and their desires and generating a solution after a solution. This is time-consuming and can be irritating but when the solution is found, the fundamental problem and related problems are solved at the same time, and the client is satisfied. [16]

In the near history, IT systems were designed and used by IT professionals only, and usability testing was not in a key role. The systems were intended for peers that had similar skill set as developers. Today, however, an increasing number of IT products are designed for the wider audience, and more work related to usability and ease of use needs to be done. Guiding the design process, for example, usage and task profiles can be used. The generic usage profiles are (1) novice or first-time users, (2) knowledgeable intermitted users and (3) expert frequent users. Task profiles are a set of tasks (high- or middle-level) that the user is expected to do with the IT artifact and can be used as a guide in the design process. [160]

The established development process of an IT artefact requires user-interface architects, designers, standardized components, specification methods (e.g., universal language), menu trees, state charts, transition diagrams, user action notation and automated tools (e.g., programming toolkits, prototyping tools). Standardization and consistency within systems are required to avoid even dangerous errors. [160] The quality of design is a safety issue as well. For example, industrial accidents caused by human error are often fundamentally design problems, and a result of inadequate design and development process. [16] To improve the quality of a product, things to pay attention to are, for example, consistent terminology and spelling. People get frustrated with complexity and defective design confuses and get people even angry. Clever design avoids the frustration and struggle of users. [16]

People have different qualities that are essential to consider. The qualities are physical, cognitive, motoric, perceptual, or related to personality, culture, disabilities, or age. For example, people with limited color-vision appropriate colors need to be chosen for *user interfaces* (UI) to enable effortless usage. Depending on the repeatability level of the task, a simplicity-level of actions must be adjusted. The more often the task is needed, the simpler the set of commands is required. [160] In *short-term memory*, working memory can be stored in from five to seven elements which is good to keep in mind with design. It takes time and effort to relocate the information into *long-term memory* and use information stored there. [16] Design is also related to time consumption. For example, the design of an interface of a game must be done with care since the time required in familiarizing application shortens the time of the actual gameplay [20].

Carey (1988) has summarized that good design is reliable, safe, standard, effective, inexpensive and built on schedule [192]. However, fulfilling all these requirements can be a challenge. When the design process is well executed, products are easy and pleasurable to use. Good design does not expect users to understand a machine or read long instructions, instead it communicates rules to follow, understands human behavior, and utilizes these as a base in the design process. It is important to highlight that everything is not possible since machines are involved. The functions available are limited, and the rules machines obey are quite simple. [16] Good digital design is a balance between human and machine qualities. This issue is studied more closely in the following, and ICT related design principles where people are in focus are introduced.

2.3.1 Interaction design and behavioral science

Three main design trends related to the design of IT artefacts are *industrial design*, *interaction design* (IxD), and *experience design*. The industrial design optimizes the value, appearance, and function of the product to serve both the user and manufacturer. In experience design, the interest lies in the quality and enjoyment of the experience. Design and especially IxD is one of the three main components of this study. IxD concentrates on enhancing the understanding of how people use products and elements that communicate and keep the user informed of the system's processes. [16] IxD designs acts that are expected to happen in other words, the intended use of products. Balancing between the human capabilities and system qualities where functions are utilized in the best possible way is a challenge of IxD. [193] [194]

IxD (i.e., *HCI design discipline*, *User centered design*, *Human-centered design* (HCD)) is interested in human interactions with digital items. It relates to *user interface* (UI) and *user experience* (UX). In HCI studies, the special interest is on what impacts computers have on human activities and work. [195] [58] [196] IxD is strongly connected to social and behavioral sciences and has developed inside these traditions. Behavioral science is interested in human behavior and has four goals: (1) to describe, (2) to predict, (3) to discover the cause and (4) to explain human behavior. [58] The cause considers temporal presence, for example, the chronological order, the covariation of cause and effect, and the possibility of alternative explanations [197]. According to Hallnäs and Redström (2006), the connection of IxD (e.g., methods and thinking) to empirical sciences is light since defining use and users is a struggle. [194]

IxD understands both psychology and technology. In the design process, the needs of humans come first, and human capabilities and behavior are considered. The aim of the design is to use human affordance in the best way and adapt the design to fulfill requirements. Keeping the user informed about what can be done (the Gulf of Execution), what is happening (the Gulf of Evaluation), and what will happen next is essential. The situations where the user does something unexpected with application, the importance of communication especially rises, and the quality of the design is tested. Defining specifications, testing continuously (rapid tests), observing, and making modifications is the ideal workflow of the design process. [16]

According to Adams (2014), for example, the game interface should provide the user with the information needed to proceed. For example, where user is, what and how user is doing, what should be done next, what are challenges user is facing, does user have all the components that are needed to succeed in the game, whether actions made were successful or not and is user in the danger of losing or in the way of victory. The information provided should preferably be in graphical or symbolic form. Audio, as well can be used to inform the user. [198] The volume of choices offered for the user should be right, not too many but enough. Otherwise, the user can get overwhelmed or frustrated. [199]

The special qualities of VR need to be taken into account when VR is the design platform. In VR, it is novel that the user can freely choose where to go or what direction to watch and therefore there is no need to design angles. User freedom can be challenging in terms of design. User can end up watching something that is not intended to, and on the other hand ignore relevant information. More scaffolding as normal is needed and more attention paid on the design of the environment. As an inspiration can be used traditional theatre and modern cinematography to find elements relevant. Telling a story in VR can be used staging, movement, light, spatial sound, focus, and depth of field (DOF). Longer takes rather than quick cuts are recommended to use. [200] [178]

Participatory design (PD) can be used to involve people who will be using the product in the design process. PD is not often combined with the IT artifact design, but it includes principles that can be used as an inspiration for the process. Implementing IxD without involving users is not possible and between IxD and PD is a strong connection. The collective design is executed with people instead of for them. [17] Utilizing the available knowhow of the team in the best possible way is the aim [160]. *The Bauhaus, Scandinavian and Nordic designs* are included in the family of PD and disagree on conditions where special experts have unchallenged power and authority on design. [17] [201]

In Scandinavian participatory design democracy, democratization, open discussion of values, expectations, diversity of opinions, and contradictions are seen as valuable elements for the process [202]. Better products (more knowledge involved), realistic expectations of users, the decreased resistance of a change and increased democracy in the environment (i.e., workplace) are benefits of a collective design process [203] [204]. In the collective resource approach, the democratic result means that every member has a voice, all voices are heard, and they have an impact on the outcome [205]. This hybrid network of minds has its origin in the 1970s [201]. According to literacy, the principles of PD have been used, for example, designing digital music games for children and children have contributed with fresh ideas and improved the product [206]. In this study is adopted principles from PD alongside with IxD.

2.3.2 Principles and models for the design path

Design has no precise path, and new goals can rise during the constantly changing process. Every design project is different and dependent on multiple factors. It is difficult to predict, time and a budget needed for the process beforehand. [160] The drivers of the design process can vary, and choosing which path and model to follow in the process is important. The process can be done, for example, by technology, by competition, or by aesthetic driven. [16] The drives of the process are often visible in the final product, which is something that is worth considering when the model for the design process is selected [207]. When it comes to learning, pedagogy driven process is expected [134]. Jeffrey and Chrisnell (2008) delivered five reasons why products can be difficult to use: (1) development has focused on the machine or system, (2) target audience have changed and adapted, (3) designing usable products is difficult, (4) team specialist did not work in integrated ways or (5) design and implementation did not match. [56]

Various methodologies are developed to ease the software development process such as the *Technology Acceptance Model* (TAM), *GEM, SCI, the Contextual Game Experience Model*, and the *Contextual Model of Learning* [160]. In this study, the focus is on design models and frameworks that can be used as a guideline of the process. The models introduced are *the double-diamond model, the three pillars of design*, *Norman's seven stages of action model (SSA), methodology based on an interactive screenplay, lifecycle for interactive-systems development* (LISD)*,* and *Pedagogy-Space-Technology framework* (PST).

The first and simplest model introduced here is Norman's double-diamond model, and it can be utilized, for example, in the design thinking process. It consists of two diamond cycles. The first cycle is for defining the problem and the second cycle is to identify the solution for the problem. In both cycles, the number of possible problems or solutions increases until in the convergence phase only one result will be left (i.e., divergence-convergence process). [16] The second presented model is the three pillars of design. The model has design layers that interact with each other. The layers include (1) guideline documents, (2) UI management system and rapid prototyping tools, and (3) usability laboratories and iterative testing. The idea in guideline documents is to provide a set of guidelines to follow to gain harmony in the design process. The foundation of the three pillars is, firstly, theories and models of HCI and secondly experimental research are. In the HCD process cycle, phases of observation, idea generation, prototyping and testing are following each other. [160] [16]

Testing and getting feedback are important parts of the design process. Validating the product prototyping, usability and acceptance tests are needed. Iterative testing is a recommended method enabling the early modifications of the product based on the received feedback. Testing can be done in usability laboratories where all actions made are carefully monitored. Surveys, interviews, and group discussions can be used to recognize users' needs, perception and experience. In acceptance tests, the program is evaluated through measurable preset goals regarding the performance of hardware and qualities of software. If the system fails a test, modifications need to be made before a new test is performed. [160]

Guidelines on how to build a good HCI have been published based on the experience received of usage of IT systems, for example, by IBM. The downside of these documents is that they can be lengthy and thus not always easy to use. [208] Jeffrey and Chrisnell (2008) introduced ten techniques to use to develop usability: (1) ethnographic research (i.e., observing users e.g., to build user personas and scenarios), (2) participatory design, (3) focus group research, (4) surveys, (5) walk-throughs (e.g., imaging the path user are expected to follow with an early prototype), (6) open and closed card sorting (e.g., asking participants to sort or name categories related to content of the website), (7) paper prototyping, (8) expert or heuristic evaluation (i.e., a list of statements to be checked usually by professionals. A Professional can also be a “double” specialist and have knowledge of the area involved (e.g., education as well), (9) usability testing (i.e., user testing) and (10) follow-up studies (i.e., studied after the initial release). [97] A follow-up study is the practice recommended. The time before performing follow-up study should be from two weeks up to one year. In case the learned content is soon forgotten, the educational gain stay minor. The biggest benefits of follow-up study are gained with short interventions. [56]

Nielsen’s heuristics are commonly used in usability assessment of UI. The Nielsen’s heuristics include 10 factors: (1) visibility and status, (2) match between the system and real world, (3) user control and freedom, (4) consistency and standards, (5) error prevention, (6) recognition rather than recall, (7) flexibility and efficiency of use, (8) aesthetic and minimalistic design, (9) well-structured features and (10) use of default values. Each factor includes statements that are evaluated on scale 0 to 5. Value five represents a serious usability problem (i.e., usability catastrophe) and value 0 not a usability problem at all (0 = not a problem, 1 =cosmetic, 2 = minor, 3 = major, 4 = catastrophe). [208] [209] [210] [211] The eight golden rules of dialogue design include similar principles than Nielsen’s heuristics. Rules are (1) strive for consistency, (2) enable frequent users to use shortcuts, (3) offer informative feedback, (4) design dialogues with closure, (5) offer a simple error handling, (6) permit an easy reversal of actions, (7) support an internal locus of control (i.e., user is in charge) and (8) reduce a short-term memory load. [160]

Testing is also important in the third model introduced. Norman’s seven stages of action (SSA) cycle provide a tool to better understand human actions. Stages of the model are (1) goal, (2) plan, (3) specify, (4) perform, (5) perceive, (6) interpret and (7) compare. [16] Seven fundamental principles of design have been developed based on the SSA model, that can be useful in the process of testing an application. The principles include similar features than Nielsen’s heuristics [209]. (1) Discoverability refers to informing the user about what is possible to do and what is currently happening with the system. (2) Feedback is something that lets the user know that the request is noticed, and the system is processing it. The right amount of feedback and prioritizing is important. (3) The conceptual model describes simplistically how something is functioning. Model is the appearance, the feel and functions of the product built. The system image includes the physical structure, and for example, documentation. The users’ (mental) model is an image based on personal interactions with the product and knowledge of the system image. (4) Affordances inform what can be done with an object. For example, a chair affords to sit. (5) Signifiers are indicators (e.g., mark and sound) of correct behavior. (6) Mapping specifies, for example, which switch is connected to which light of the ceiling. (7) Constraints (physical, semantic, cultural, logical) need to be considered in the design process. [16]

Building a high-quality system is a challenge and Shneiderman (1992) represented the LISD (the fourth model) which can be used as a framework. The stages of lifecycle are (1) collect information, (2) define requirements and semantics, (3) design a syntax and support facilities, (4) specify physical devices, (5) develop a software, (6) integrate a system and deliver to users, (7) nurture the user community and (8) prepare an evolutionary plan. [160] This model is the first that also considers actions after the delivery which are very important as well. However, many models are cyclic and can be considered continuing after release.

Prieto et al. (2015) designed a methodology based on an interactive screenplay that can be used in educational purposes. It follows the principles of PD but offers a pattern for the process. The methodology is based on narrative, which is seen advantageous in terms of collaboration, especially when non-technical members (e.g., educators, artists, and writers) are involved. This approach has different abstraction levels from detailed to broad to serve the needs of the multidisciplinary team. For example, the game designed is divided into chapters and scenes. The model consists of pre-phases and phases where educational challenges, type of the game, a story and characters, chapters and scenes, emotions, adaptation, and collaboration are designed. [212]

The PST framework that was mentioned earlier in the chapter related to education follows the goals of PD as well and includes lifecycles with two focuses (1) Conception and Design and (2) Implementation and Operation. As a pedagogical aspect is considered, for example, *"What type(s) of teaching are we trying to foster and why?"*. Questions related to space is, for example, *"Which aspects of the space design and equipment worked, and which did not and why?"*. In the third element technology are considered, for example, *"What pedagogical improvements are suggested by the technology and what were the unexpected impacts (positive and negative) of the technology on learning and teaching?"*. By following PST framework and set of questions is expected to produce better learning spaces. [134]

3 Rationale of the study

3.1 Research plan

Research about learning materials has been done in general and digital learning is studied in an educational context. However, among VR learning materials the research done is lacking [213]. It is essential to receive detailed information of features important to include in VR learning materials before the technology is ready to be implemented into education. With the aid of the results received from this study, better and more accurate learning materials can be designed into VR environments. The aim of this study is to name the key components of valuable VR learning material.

In this study, is utilized Bundsgaard and Hansen (2011) *"a holistic framework for evaluating learning materials and designs for learning"* model. The framework is based on previous research findings that use different methods, for example, systematic tools to find certain parameters, investigate reasons why teachers select certain materials for teaching, utilize surveys, evaluations, testing materials with teachers and students and, observe learning impacts. Bundsgaard and Hansen (2011) defined a design for learning to be*: "a constellation of artifacts (which can be called learning materials) arranged (in space) and articulated (in time) by someone with the intention to initiate and support someone's learning"*. A holistic evaluation of learning materials includes (a) *the potential of learning potential* (i.e., learning potential of a text), (b) *the actualized learning potential* (i.e., learning potential in the classroom context) and, (c) *the actual learning* (i.e., impacts on students' competences). The holistic evaluation can be performed from four different perspectives: students', teachers', situational or society's perspective. Situational perspective includes the environmental factors (e.g., time, space, participants). Society on the other hand provides culture, law, and national curriculum related perspectives. [99]

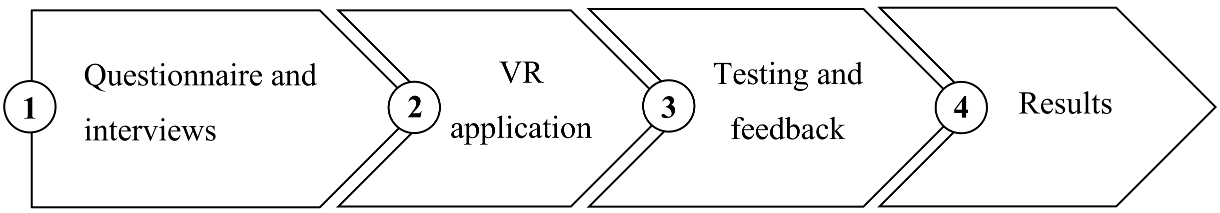


Figure 7. The structure of the study includes four phases.

In this study, it is decided to strongly include the perspectives of both teachers and students in the research plan, but situational and society's perspectives are considered as well. The basic structure of the study consists of four phases (1) questionnaire and interviews for teachers, (2) building a VR application based on the result of the phase 1, (3) testing the application with students and gathering feedback and finally (4) analyzing the data and reporting results (see Figure 7).

This study also follows the principles of DSR paradigm [63] (see Figure 8) which consist of three components: knowledge base, the environment and design. Models, frameworks and methodologies together form the knowledge base. In the study, is utilized a combination of methodologies of survey research (questionnaire, interview), an exploratory user testing, and data analysis techniques. The research environment of DSR includes teachers, students, and the school environment. The technology utilized is VR. Education and pedagogy form the organizational level of the DSR paradigm and understanding the school environment and educational needs are essential. In the design phase an IT artifact is developed: a VR application, and it is evaluated and tested in a case study set up in an authentic classroom context. The collected data is mainly quantitative. The data is analyzed using statistical analysis, and results are discussed later in the study. The quality of the system is measured with quantitative quality metrics to determine what level the system possess quality attributes and achieves user needs and expectations [14]. ISO/IEC 25000 standard (SQuaRE i.e., System and Software Quality Requirements and Evaluation) is developed to standardize the process of quality assurance of software development [214]. The ideas of ISO 25000 standards, are utilized as part of the study.

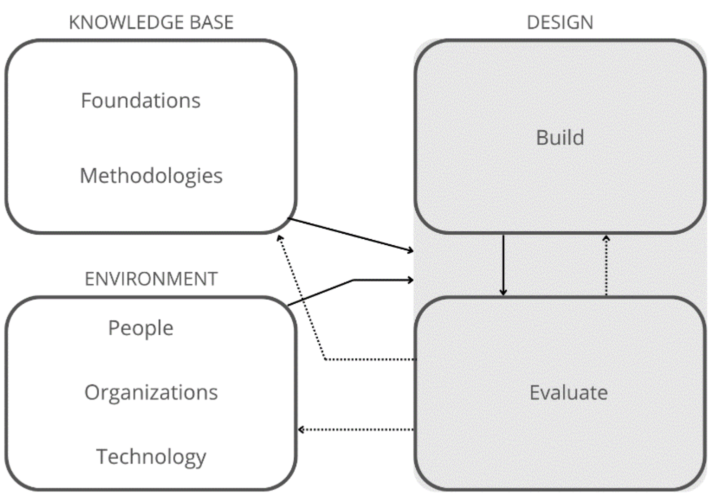


Figure 8. An adapted DSR framework from the original Hevner et al. model from 2004 [62].

In this study, as the foundations of DSR knowledge base are combined ideas of multiple models and frameworks as shown in Figure 9. The core of the combined design model of the study is formed by the concepts of PST framework and is framed with the HCD process cycle and the first five stages of LISD model. The space of PST framework is in this case very much dependent on technology since a VR space, a virtual environment is involved. However, it is important to consider the constraints of the physical space as well, and limitations of the classroom and the school environment. The ideas of participatory design approach are utilized as well, and teachers and students are involved in the design process through questionnaires, interviews and the user testing.

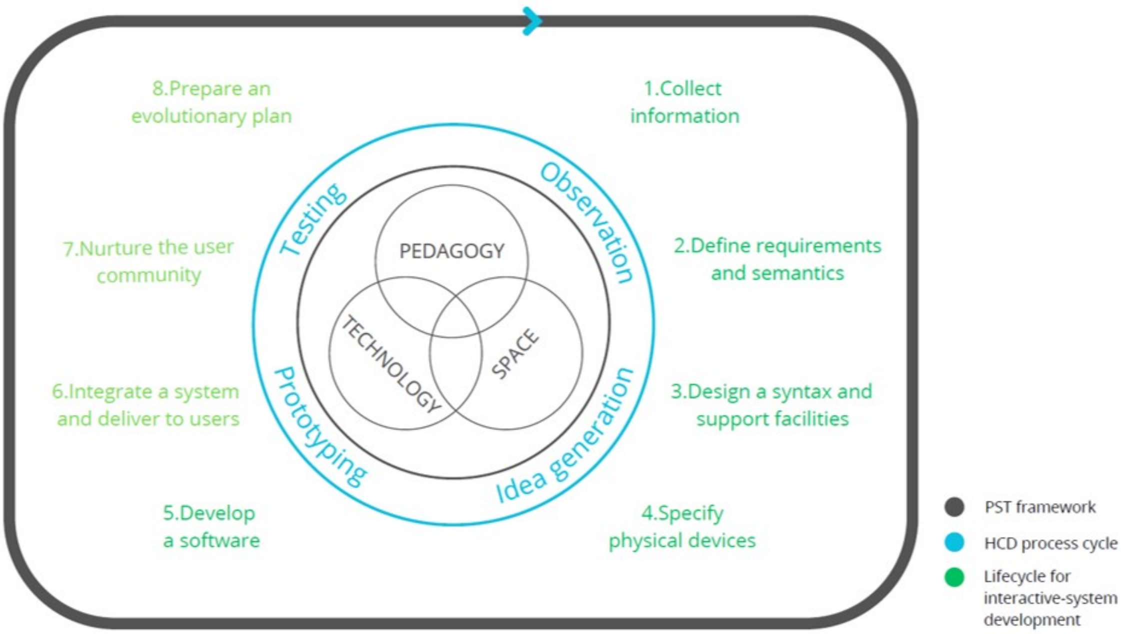


Figure 9. The combined design model of the study

In the first stage of LISD, collect information, the former research done in the area is familiarized, and new information collected. The former scientific knowledge was introduced earlier, in the chapter on background. In this study a special interest relates to qualities characteristic to biology, the subject the material is designed for, since each subject has unique features important to consider in the learning material design. In the second stage of LISD, the collected data is analyzed and defined requirements. After that, the device and syntax are specified. VR application, in other words the learning material is developed in the fifth and final stage of the LISD, based on the results and requirements defined. The age group studied is lower secondary school students since manufacturers recommend VR headsets to be used by not younger than 13 years old [188]. The VR learning application is tested with students (user testing) in an authentic classroom situation and feedback is collected in the form of a questionnaire. Students are as well observed during the testing sessions to trace the learning experience and the quality of the learning material.

3.2 Research questions and statistical analysis

The six research questions that the study aims to answer are:

RQ 1. What are the main components of valuable VR learning material?
RQ 2. What kind of role are teachers preferred to have during the VR session?
RQ 3. Does teaching experience have an impact on preferred VR features?
RQ 4. Is there a difference between teacher's and student's VR preferences?
RQ 5. How can VR learning be implemented into education?
RQ 6. What are the challenges of VR learning?

The collected data and statistical analysis are used to solve research questions. Statistical significance (P) is analyzed with MANOVA. With Mann-Whitney U-test the teachers' and students' preferences are compared. Statistical analysis, Cronbach alpha and descriptive statistics are calculated with SPSS program and R values are calculated with Excel program.

Work plan and schedule:

Literature search: November 2022 to December 2025

Questionnaire and interviews for teachers: January to March 2023

Development of a VR application: February to May 2023

User testing and feedback questionnaire for students: May 2023

Writing the Master Thesis: May 2023 to December 2025

4 Teachers' questionnaire

Survey research is a process to receive information about people's attitudes, beliefs, facts, demographics, and behaviors [215]. Interviews and questionnaires can be used as tools for collecting data [189]. In phase one of this study, (see Figure 7) a questionnaire to identify what are the features teachers prefer the most in the VR learning materials is delivered. The questionnaire is addressed to biology teachers and student teachers. Interviews are to complement answers received from the questionnaire, and to get more details. In the interview the same structured format as in online questionnaires is used, however, the Zoom video conference platform is utilized for personal interview, and the form is filled in by the interviewer (*Computer-Assisted Video Interviewing* (CAVI) [216] [217]).

4.1 Methods and data collection from teachers

4.1.1 The Questionnaire form design

Designing a questionnaire and choosing questions to include is a process that needs to be done with care. The most important issue is that the questions of the questionnaire need to be bound to research questions. Asking irrelevant questions for participants may result in participants terminating the survey in an early stage causing a lower response rate. Certain types of survey are discovered to have a higher response rate, for example, telephone surveys. In this survey an online questionnaire is used which is a convenient and relatively common method to collect data. A good questionnaire is simple, avoids double-barreled, loaded questions and negative wording as well as constant agreement. The questionnaire can include closed- and open-ended questions. In closed-ended questionnaires, the scale can be from choosing simply "yes or no" to 5-point or 7-point scale and labelled, for example, from strongly agree to strongly disagree. Aiming at a balanced scale is a good practice. [58] The rating scale can be graphic, semantic differential or non-verbal (e.g., symbols) and the frequency high or low depending on the needs of the study [218].

The teachers' online questionnaire (see Appendix 1) was conducted on the Webropol survey and reporting platform. The platform is an online platform and can be answered through, for example, a web-link, e-mail, QR code, or SMS [219] and many different question types are enabled by the platform. This study utilizes a questionnaire with multiple-choice, open-ended, Likert scale, dropdown, and demographic questions.

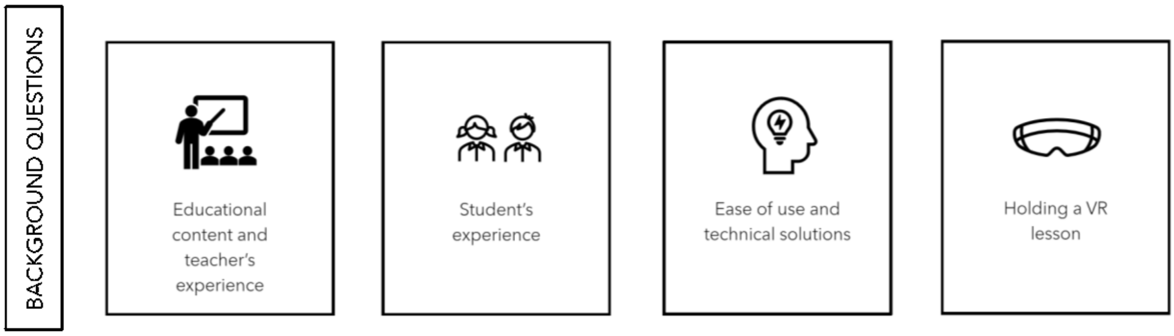


Figure 10. Themes of question sets of the teachers' questionnaire.

Placing the most interesting and relevant questions first is a way of drawing people's interest and receiving more answers [58]. Grouping questions under themes makes the appearance of the questionnaire more professional and results in high-quality replies [220]. In this study, the questions are grouped in five main themes, and the relevance of themes is considered in the placing, however (A) background questions are placed at the very beginning of the questionnaire as illustrated in Figure 10. Background questions are to recognize, for example, teaching and VR experiences. In the second theme, questions relate to (B) educational content and teacher's experience. The third set of questions are about (C) the student's experience. The fourth theme is about (D) ease of use and technical solutions. The final theme includes questions related to (E) holding a VR lesson. In the very end of the questionnaire is an open-ended question to provide an opportunity to give feedback and to comment on the questionnaire and the study in general. The main themes (A–E) with more details and purposes are introduced next.

(A) Introduction and background questions part is at the beginning of questionnaire and provides a brief introduction about the purpose of the survey, who it is for, how the data is handled and stored and who can be contacted for further information in case needed. It is important to deliver this information since, response sets, for example social desirability, can cause bias for dataset and guide people's responses. The bias can be avoided by openly communicating the aims of the study, providing confidentiality and feedback as a form of results. [218] The background questions clarify gender, age group, teaching experience, the confidence level of using digital learning material and the knowledge related to VR (with a short definition what VR is about). After that it is discovered the teachers' experience related to VR, the usage of VR with students, how many headsets have been used and what models and applications. At the end of question set is an open-ended question to provide an opportunity to be more precise and to offer more information about details related to the experience.

(B, C, D) The Likert scale question sets. The Likert scale is used with question sets (Q13) Learning content and teacher's experience, (Q15) Students experience, and (Q17) Ease of use and technical solutions as illustrated in Table 2. Sixteen statements are included in the first set of questions, and in the second and the third set have both thirteen statements. The used Likert scale is evenly distributed, five-point scale. The scale is generally accepted for measuring attitudes and the degree of agreement of participants. The statements of agreement used in the study are (1) Strongly Disagree, (2) Somewhat Disagree, (4) Somewhat Agree, (5) Strongly Agree and (3) Undecided [221]. "A somewhat agree" statement is used instead of "agree", since it has been discovered that people are willing to give rather positive than negative answers. Using "somewhat agree" and "somewhat disagree" correspond better to the value on the numeric line [222] [223] [224]. For example, if the statement of agreement would be "agree", the value should be higher than four. Kalliopuska's (2005) third neutral statement "do not disagree or agree" is renamed in the study as "Undecided" similarly as in the Likert's original set of statements ((1) Strongly Approve, (2) Approve, (3) Undecided, (4) Disapprove, (5) Strongly Disapprove) [225] and displaced last in the scale. The idea of displacing the neutral statement last is to highlight the importance of selecting between two opposite extremes "Strongly Agree" and "Strongly Disagree" and to receive fewer neutral answers.

In the fifth main theme (E) questions related to holding a VR lesson, and were to determine topics that could be beneficial to learn with VR, what is the preferred duration of the session, how often VR learning materials should be utilized in learning purposes and how many VR headsets should be available for a class of 20 students.

4.1.2 Sampling and data collection

The subject for the learning material was selected to be biology because of professional interest, and competence and practical experience. The learning application aimed to be tested in a Finnish school and for this reason the questionnaire and interviews were addressed to Finnish biology teachers and student teachers as well who have the best knowledge of the Finnish school system and the practical knowledge of the needs the educational material have related to the subject. Student teachers were included to get teachers from every stage of their careers to be involved in the survey. To have teachers with different teaching experience and background to take part was seen important to get a general opinion.

Research can be qualitative or quantitative or mixed method style which includes features of both research methodologies. Qualitative research usually collects in-depth information from quite a few people or in limited research settings. Quantitative research has a very limited area of interest but bigger sample size. [58] This study has qualities of both researches. The number of participants is quite small, which is characteristic to qualitative research. However, the data collected from questionnaires is mainly quantitative.

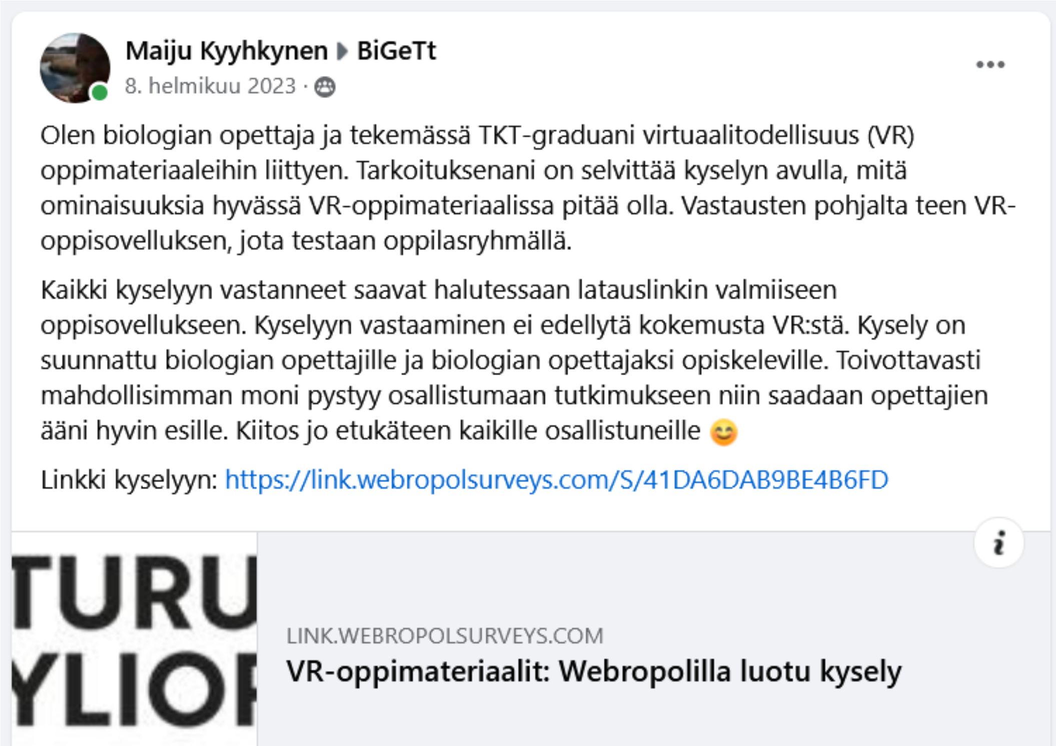


Figure 11. The invitation to take part in the survey published on Facebook explains what the purpose of the study is, who is the realizer, who can participate, what are the benefits, and why it is important to participate.

Convenience, for example, nonprobability sampling is often used to save time or for budget reasons. Students are usually used as respondents for the same reason as well. [58] In this study, a purposive sampling is used to choose participants to answer the questionnaire. Purposive sampling is a form of nonprobability sampling, and the idea is to find people who suit a predetermined criterion. [58] The responses were collected with an open link. The link with a short invitation was published in a private group on Facebook as shown in Figure 11. Facebook is an online platform (i.e., social networking site) to contact people with similar interests, to share images, videos and thoughts with an option to like and comment on other people's publications [226] The invitation to participate the study was published in a teacher's

Facebook group. The group is an online *professional learning community* for biology, geography and health education teachers (BiGeTt) and the idea behind the group is to share good educational material and to discuss education. There are published invitations to participate in surveys at times as well. The group is active with 2200 participants and delivers about four publications per week.

The aim was to receive twenty responses. The invitation to the questionnaire was published in the group on the February 8th, 2023. Since all participants of the group had at least in theory an opportunity to participate in the survey, the sampling (i.e., survey) frame of the study is 2200 people. The sampling frame is the group of people who can take part in the research [58]. In the first round ten responses were received. Follow-up reminders have proved to be an effective way to increase the response rate in previous studies [58]. In this study a follow-up reminder is used as well and sent after a week from the original on February 14th as seen in Figure 12. Finally, the questionnaire was open for responses for two weeks' time and was closed on February 21st. The total number of responses received was 15 and the final response rate was 69%. The number refers to in this case the number of people who started to fill in the questionnaire and completed it instead of all participants of the teacher's Facebook group.

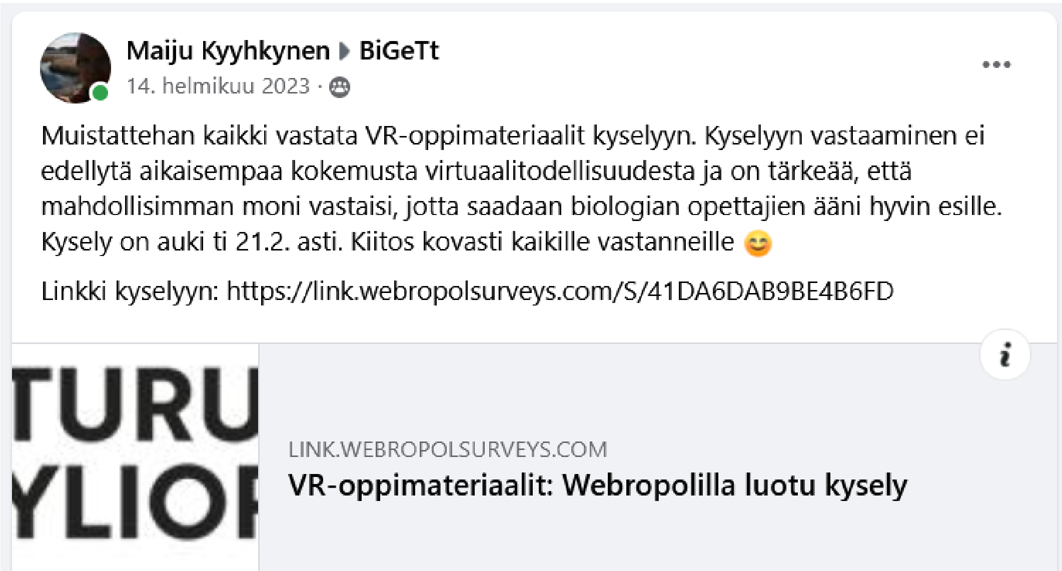


Figure 12. In the remind message it is emphasized that no earlier experience about VR is needed to participate the survey, why it is important to participate, how long the survey will still be open and finally is thanked for teachers who already have participated the survey.

Three biology teachers were also interviewed (CAVI) in this first phase of the study, to receive more detailed information. Convenience sampling was used for selecting the teachers who were interviewed. One of the teachers interviewed was a secondary school teacher, but taught at upper secondary school too. The other teacher interviewed was an upper secondary school teacher, and the third teacher educated student teachers at the university. Participating in the survey was voluntary for everyone, and the total number of participants after interviews became 18 teachers, which was considered adequate.

4.2 Analysis and results of the teachers' questionnaire

After collecting replies from biology teachers, the data needed to be analyzed to identify teachers' preferences. Combining design and statistical analysis can be difficult. However, it has been discovered that the research performed by social scientists provides guidelines. [58] [160] To discover which features to include in the VR application to be built, the collected data is analyzed with several tests. Statistical analyses are to solve research questions (RQ1 – RQ6) and to determine whether null hypotheses are true (i.e., no difference among groups) or false (i.e., a significant difference among groups) [227]. In the analysis, it is utilized mainly Likert scale question sets which were to collect data on attitudes but demographic questions and open-ended questions are utilized as well to complete reasoning. Before analysis the internal consistency (i.e., the reliability) of each theme is ensured and other required modifications for the data set made.

4.2.1 Modifications and reliability analysis of themes

The first modification made was to rearrange three themes (Q13, Q15, and Q17) with Likert scale in the teachers' questionnaire into new themes to measure more precise qualities. New themes are (1) usefulness, (2) usability and ease of use (i.e., the usage is easy and pleasant [14]), (3) safety and health, (4) teacher's role and (5) student's experience. These five themes are the themes that are discussed in the analysis phase instead of the original theme of the questionnaire. Teacher's role related statements are utilized to recognize what kind of role are teachers preferring to have in the VR lesson. This theme is used as well in the analysis to determine the priority of themes but is not used in the later examination.

In the first new theme (1) Usefulness questions relate to learning objectives and include elements of curriculum in general (C), the broad competence (L1–L7) of learning, the subject related competence (S1–S6) and the objectives of teaching (T1–T14)) [93]. Curriculum, and usefulness related questions are shown in the usefulness column of the Table 2. The second theme is (2) usability and ease of use and includes statements related to practicalities and usability heuristics (Nielsen heuristics N1–N10). Statements are visible in the usability column of the Table 2. The third theme: (3) safety and health, includes five statements as can be found in Table 2. The statements of the fourth theme measures preferred (4) teacher's role [93] [210]. Ten statements shown in the teacher's role column of the Table 2. In the fifth and final theme the statements measure (5) student's experience and are shown in the last column of Table 2.

The second modification made for the data set was to return the value 5 (undecided; "en osaa sanoa"; I do not know) in its original place in the middle of the scale (value 3). Here is a strong assumption that the respondents have meant a neutral opinion. Another option would have been to remove all these answers but since the number of teachers who participated the study (i.e., answered the questionnaire or interview) remained quite low (n = 18) [227] it would have had an impact on the results. In a small dataset, a single missing value can have a great impact on the final results, and it turned out that the number of undecided values was quite high (n = 46) and comprehended 6.2% of all answers.

The third modification that was performed was to rescale four statements the other way around. The action was required to be made since in some statements the highest value represented positive and in other negative correlation. The statements that needed to be rescaled (1 = 5, 2 = 4) measured teacher's role. Statements were 13.G., 13.H., 15.I. and 15.J., and after modifications, the data set was consistent. The higher the value, the more passive the role of a teacher. The rescaled data was used to analyze the teacher's role.

The fourth action made was to control reliability of themes: the usefulness, the usability and ease of use, the safety and health, the teacher's role and the experience. By grouping statements into themes, it was possible to increase the volume of data per condition for statistical analysis, however, the number of participants did not increase or reach the common minimum requirement of 30. The results of statistical analysis can be considered indicative and seen to support other results [56]. Cronbach's alpha indicator values (i.e., the reliability) were calculated to ensure that chosen statements of each theme were consistent, and measured the same subject matter (i.e., had internal consistency). Cronbach's alpha is a value that relates to the average of items' internal correlation coefficients and the number of items. [58] Value that is over 0.7 is normally accepted. If the α value is > 0.9 reliability is excellent, value

0.9 ≥ α > 0.8 is good, value 0.8 ≥ α > 0.6 is questionable, value 0.6 ≥ α > 0.5 is poor, and if α value is < 0.5 correlation is normally unacceptable. [228] In addition to Cronbach's alpha indicator values item-total correlation values for each statement were also calculated. Statements that scored less than 0.2 were removed from the set. Low item-total correlation values refer to problems inside theme groups. [58] In Table 1 below is presented descriptive statistics of each theme.

Table 1. Descriptive statistics. The number of selected statements of each theme group, Item-Total correlation and Cronbach's alpha values.

	N	Corrected Item-Total Correlation	Cronbach's alpha
1. Usefulness	19	0.26–0.76	0.85
2. Usability and ease of use	7	0.26–0.68	0.74
3. Safety and health	3	0.44–0.53	0.53
4. Teacher's role	6	0.29–0.66	0.73
5. Student's experience	7	0.21–0.62	0.66

19 statements were chosen to represent the usefulness theme, seven to the usability and ease of use, three statements to the safety and health, six to the teacher's role and seven to the student's experience theme as presented in Table 1. Statements that were selected to measure (1) the usefulness theme (n = 19, Cronbach's α = 0.85) can be seen in bold in the usefulness column of the Table 2. The usability and ease of use (2) statements can be found in the next column of the Table 2 (n = 7, Cronbach's α = 0.74). The theme of (3) safety and health statements are in the third column of the same table (n = 3, Cronbach's α = 0.53) and (4) the teacher's role related statements in the fourth (n = 6, Cronbach's α =0.73) and finally, statements measuring (5) the student's experience can be found in the fifth column of the Table 2 (n = 7, Cronbach's α = 0.66).

Considering Cronbach alpha values, it is noticeable that the bigger the number of statements in a theme group the bigger the measured alpha value normally is [58]. The smallest group safety and health (n = 3) has the lowest alpha value (0.53) which is a poor reliability coefficient. In student's experience group (α = 0.66) the value is also below the limit that is generally accepted. Removing two statements from the group student's experience (15.L., and 15.A.,) increases the alpha value (n = 5, Cronbach's α = 0.70). However, for this reason the quality of statements simultaneously decreases, and the two statements relevant for the group are decided to include in the set. In addition, the α value 0.66 is rather close to the value 0.7 that is acceptable. The theme safety and health is examined next. It is discovered that

Table 2. Statements of the teacher's questionnaire (Q13, Q15 and Q17) that are selected for Option A analysis are marked with A. Statements selected to Option B analysis are marked with B. Statements selected as quality requirements (R1–R12) are bold and italic. C=curriculum, L=broad-based learning (1–7), S=content area, T=goal (1–14), N=Nielsen heuristics (1–10), S=safety, H=health, A= active, P=passive, E=experience

Q13 LEARNING CONTENT AND TEACHER*S EXPERIENCE	Usefulness (C/L/S/T)	Usability and ease of use (N)	Safety and health (S/H)	Teacher's role (A/P)	Student's experience (E)
13.A. Material obeys the curriculum. (R1)	**C^{AB}**				
13.B. There is very little preparation needed for the teacher to use the material.		N		P	
13.C. Material considers diverse learners.	**C^{AB}**	**$N7^{AB}$**		**P^{AB}**	
13.D. Student gets feedback.	**C^{AB}**	**$N1^{AB}$**		**P^{AB}**	
13.E. Help is provided when needed. (R2)	**C^{AB}**	N9, N10		P	
13.F. Students learn to use VR glasses.	**$L5^{AB}$**				
13.G. The teacher can monitor students in the VR environment and see the progress.				**A^{AB}**	
13.H. The teacher is actively guiding students				**A^{AB}**	
13.I. The material intensifies the knowledge of students. (R3)	**C^{AB}**				
13.J. The material supports investigative learning.	**C^{AB}**				
13.K. It is easier for students to understand abstract concepts.	T7				
13.L. Student learns causality.	**L1, T7, T3, $T6^{AB}$**				
13.M. Student learns methods to collect biological data.	**C, L4, S1, T1, T2, $T8^{AB}$**				
13.N. Student learns to apply biological knowledge.	**C, L4, $T8^{AB}$**				
13.O. The application provides a calm learning environment.	**C^{AB}**	**$N8^{AB}$**	**H^{A}**	P	
13.P. Teacher receives positive experience.					E
Q15 STUDENT'S EXPERIENCE					
15.A. Student experience things instead of standing by and watch. (R4)	C				**E^{AB}**
15.B. Student can explore the environment. (R5)	C, L1, S1, S2,				**E^{A}**
15.C. Student can use creativity	**L1, $L2^{AB}$**				**E^{A}**
15.D. Student gains new experiences. (R6)	**C^{AB}**				E
15.E. Student operate independently in the VR environment.	C, T8				E
15.F. Student can proceed at one's own pace.	C	N3		P	E
15.G. Application encourages students to act and try things.	**C, $L7^{AB}$**				E
15.H. Students can collaborate, discuss and find solutions together	**C, $L2^{AB}$**				**E^{A}**
15.I. Students can see each other in the virtual environment	**$L2^{AB}$**			**A^{AB}**	**E^{A}**
15.J. Student can see the teacher in the virtual environment.				**A^{AB}**	**E^{A}**
15.K. The motivation to learn the subject increases. (R7)	**C^{AB}**				E
15.L. Student can experience something that is not possible in real life e.g., fly. (R8)					**E^{A}**
15.M. Student will get positive learning experience. (R9)	**C, $T14^{AB}$**				E
Q17 EASE OF USE AND TECHNICAL SOLUTIONS					
17.A. Application is easy to use. (R10)		N		**P^{AB}**	
17.B. There is a tutorial that guides when using at first time.		**N9, $N5^{AB}$**			
17.C. Student's access to other apps is restricted.		N8		**P^{AB}**	
17.D. The appearance is esthetically pleasant.		N8			
17.E. There can only be seen functions that are currently needed.		**$N8^{AB}$**			
17.F. The environment resembles the real world.		**N2, $N4^{AB}$**			
17.G. It is safe to use. (R11)	**C^{AB}**		**S^{AB}**		
17.H. Moving does not cause motion sickness. (R12)			**H^{A}**		
17.I. The usage of the application doesn't expose students to game addiction.			H		
17.J. No unnecessary data is collected from users			S		
17.K. The application is used with touch controllers					
17.L. Application is used with hand movements.		**N2, $N4^{AB}$**			
17.M. There are few touch controller buttons or gestures to learn.		N8			

removing statements does not increase the alpha value of the theme but other way around it decreases. In the safety and health theme it is decided to include all the original three statements.

Later in the study, it is examined whether these lower alpha values have an impact on the results. The comparison is made by running two parallel analyses (Option A and Option B). The option A analysis includes the statements discussed above, but option B includes one statement from both themes only. The theme safety and health include 17.G." It is safe to use" and the student's experience theme includes statement 15.A." Student experience things instead of standing by and watching".

4.2.2 Choosing features to be implemented

The comparison of different themes aimed to answer two research questions RQ 1: "What are the main components of valuable VR learning material?" and RQ 2: "What kind of role are teachers preferring to have during the VR session? ". After the reliability of each theme was ensured, the following procedure was to calculate the average variables for each group. These numbers were to analyze whether there is a significant difference between themes.

In this study the MANOVA analysis is utilized (i.e., the repeated measures ANOVA procedure in the SPSS) since multiple measures of same participants with same scale can be compared with each other. One-Way ANOVA is for analyzing variances, but there can only be one dependent and one independent variable. In General-Linear-Models multiple independent variables are allowed. [58] [228] Related to the RQ 1 it is examined whether there is a significant difference between different themes and whether one theme is possibly more important to include into the learning material than other. The average levels of the different sum variables are compared with each other. The same scale has been used for each statement. In addition, it is used 95% confidence interval and Bonferroni's adjustment for multiple comparisons as post hoc analyses. All five themes listed in the columns of Table 2 i.e., variable groups (Within-Subject Variables (same participants in each measure)) are examined and pairwise compared with each other to identify significant differences. Greenhouse-Geyser estimates of sphericity ($epsilon^A = 0.58$, $epsilon^B = 0.52$) which corrects degrees of freedom is used since Mauchly's test indicates violation of the assumption of sphericity. It is made two parallel runs A and B and found a significant difference in both cases and the null hypothesis is rejected. $F^A (2.33, 39.60) = 24.50$, $p^A < 0.001$, ($\eta^2 = 0.59$) and $F^B (2.06, 35{,}01) = 52.78$, $p^B < 0.001$, ($\eta^2 = 0.76$).

Table 3. Descriptive statistics and significance values of MANOVA analysis of two optional A and B cases. * = The significant mean difference compared to other groups, 1 = except the usefulness, 2 = except the usability and ease of use, 3 = except the safety and health, 5 = except the student's experience.

	N	Mean^A	Std. Deviation^A	p^A	Mean^B	Std. Deviation^B	p^B
Usefulness	18	4.25	.43	**<.001*[2,3]**–.82	4.25	.43	**<.001*[2]**–.07
Usability and ease…	18	3.93	.70	**<.001*[1,5]**–1.00	3.93	.70	**<.001*[1]**–.07
Safety and health	18	4.44	.47	**<.001*[1]**–.82	4.94	.24	**<.001*5**–1.00
Teacher's role	18	3.29	.74	**<.001***	3.29	.74	**<.001***
Students experience	18	3.90	.49	**<.001*[2]**–1.00	4.94	.24	**<.001*[3]**–1.00

* The significant mean difference compared to other groups, [1] except usefulness, [2] except usability and ease of use, [3] except safety and health, [5] except student's experience. A = option A, B = option B.

In the first (option A) MANOVA analysis teacher's role is found the only theme that differs from all other four themes significantly ($p < 0,001$) as listed in Table 3. Student's experience and safety and health themes differ significantly from each other ($p = 0.018$) and other themes have similarities with one or two other themes as shown in Figure 13.

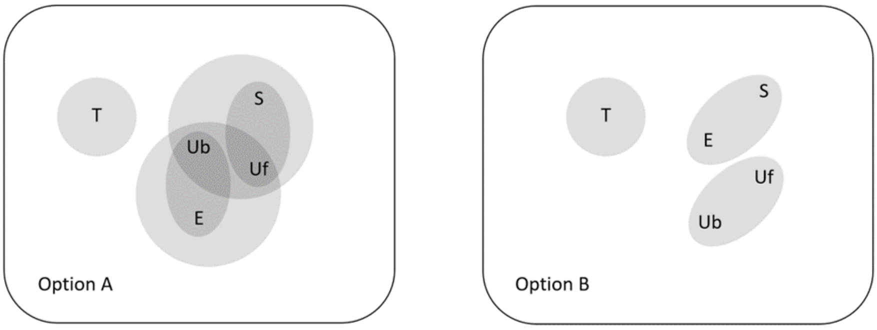


Figure 13. The similarity and overlapping of themes (variable blocks) of the teachers' responses found from MANOVA analysis. T = the teacher's role, Ub = the usability and ease of use, E = the student's experience, S = the safety and health, Uf = the usefulness

Passive teacher's role has the lowest mean value ($M^A = 3.29$, $SD^A = 0.74$) as shown in Figure 14. Student's experience ($M^A = 3.90$, $SD^A = 0.49$) is considered as important as usability and ease of use ($M^A = 3.93$, $SD^A = 0.70$). Safety and health ($M^A = 4.44$, $SD^A = 0.47$) as important as usefulness ($M^A = 4.25$, $SD^A = 0.43$). Usefulness, usability and ease of use and safety and health are considered all equally important as well as usability, usefulness and student's experience.

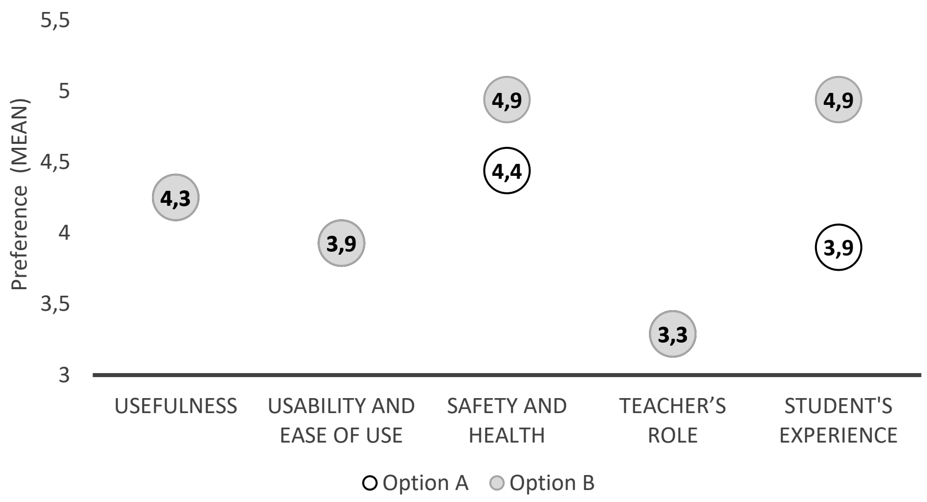


Figure 14. Comparison of mean values of Option A and Option B

Similarly, as in the first run, in the parallel analysis run, option B, the teacher's role is found the only theme that has significant difference from all other groups ($p^B < 0{,}001$) as shown in Table 3. The student's experience group is overlapping with safety and health and usefulness is overlapping with usability and ease of use as shown in Figure 13 and Figure 14. Teacher's role has the lowest mean value ($M^B = 3.29$, $SD^B = 0.74$). The student's experience is considered as important as safety and health ($M^B = 4.94$, $SD^B = 0.47$). Usefulness ($M^B = 4.25$, $SD^B = 0.43$) and usability and ease of use ($M^B = 3.93$, $SD^B = 0.70$) are considered equally important.

In conclusion, MANOVA analysis reveals that in both options A and B the teacher's role (T) is the only theme that differs every other theme significantly (see Figure 14). Themes are partially overlapping with each other, for example, since less teacher involvement is needed with better usability and this enables more passive role of teacher. Four statements from the Usability theme include also in the teachers' role category which is 50 % of the statements of the theme and similarity of these two themes is evident. Overall, the results of these two runs (A and B) are also controversial since in the option A the safety and health theme and student's experience theme are significantly different and in option B considered similar.

The idea of the comparison options A and B was to determine whether lower alpha values that were earlier accepted (S/H and E themes) had an impact on results. It appears that there can be an impact. Finally, however, it is decided to choose option A, for further analyses since it represents better the data set as a whole and teachers' preferences because more statements are involved. In addition, although only few statements were selected in option B to represent themes with low reliability values (i.e., safety and health and student's experience) there were no success in distinguishing groups (i.e., no significant difference found). Similarities and overlaps were found between the groups in both cases A and B.

Choosing statements to be implemented – the new categorization

Since it could not be found statistical difference between themes earlier, neither with option A or B, through MANOVA analysis, and all themes were discovered very much equally important, to answer the research question RQ 1: "What are the main components of valuable VR learning material?", it was decided to compare the average levels of individual statements to select which features to include in the application. Statements were divided into four categories to enable closer examination: very important ($M \geq 4.5$), important ($3.5 \leq M < 4.5$), undecided ($2.5 \leq M < 3.5$) and least preferred features ($1.5 \leq M < 2.5$) based on teachers' preferences.

Generally, functional and non-functional requirements exist. Functional requirements are tasks that system or component is expected to perform [14]. Non-functional requirements (i.e., quality requirements) include reliability, compatibility, functional suitability, flexibility, maintainability, performance efficiency, interaction capability, safety and security [229] [230]. In this study, usefulness and usability aspects are incorporated into non-functional requirements as well.

Table 4. Descriptive statistics of the three main themes: Q13 Learning content and teacher's experience, Q15 Student's experience and Q17 Ease of use and technical solutions.

	13.A.	13.B.	13.C.	13.D.	13.E.	13.F.	13.G.	13.H.	13.I.	13.J.	13.K.	13.L.	13.M.	13.N.	13.O.	13.P.
Min	4	2	1	1	3	1	2	1	4	2	4	2	2	2	2	2
Max	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Mean	**4,56**	4,28	3,94	4,00	**4,50**	3,94	4,00	3,06	**4,83**	4,11	4,39	4,44	3,89	4,11	3,83	3,72
Median	5	4,5	4	4	5	4	4	3	5	4	4	5	4	4	4	4
Std. Deviation	0,51	0,96	1,21	1,14	0,62	1,30	1,03	1,26	0,38	1,02	0,50	0,78	1,02	0,83	0,99	0,89

	15.A.	15.B.	15.C.	15.D.	15.E.	15.F.	15.G.	15.H.	15.I.	15.J.	15.K.	15.L.	15.M.
Min	4	4	2	4	3	4	4	2	1	1	4	2	4
Max	5	5	5	5	5	5	5	5	5	4	5	5	5
Mean	**4,94**	**4,89**	4,17	**4,83**	4,39	**4,72**	**4,78**	4,11	2,39	2,33	**4,61**	**4,50**	**4,83**
Median	5	5	4	5	4	5	5	4	2	2	5	5	5
Std. Deviation	0,24	0,32	0,92	0,38	0,61	0,46	0,43	0,83	1,14	1,03	0,50	1,04	0,38

	17.A.	17.B.	17.C.	17.D.	17.E.	17.F.	17.G.	17.H.	17.I.	17.J.	17.K.	17.L.	17.M.
Min	3	2	3	2	2	1	4	4	1	4	1	2	2
Max	5	5	5	5	5	5	5	5	5	5	4	5	5
Mean	**4,78**	4,44	4,44	4,17	4,33	3,56	**4,94**	**4,56**	4,00	**4,89**	2,94	3,39	4,06
Median	5	5	4,5	5	4,5	4	5	5	4	5	3	3	4
Std. Deviation	0,55	1,04	0,62	1,25	0,84	1,54	0,24	0,51	1,24	0,32	0,87	0,92	0,94

Table 5. The statements evaluated “very important” in the teachers’ questionnaire (M ≥ 4.5). The quality requirements R1–R12.

R1 Material obeys the curriculum (13.A.).
R2 Help is provided when needed (13.E.).
R3 Material intensifies the knowledge of students (13.I.).
R4 Student experience things instead of standing by and watching (15.A.).
R5 Student can explore the environment (15.B.).
R6 Student gains new experiences (15.D.).
R7 The motivation to learn the subject increases (15.K.).
R8 Student can experience something that is not possible in real life e.g., fly (15.L.).
R9 Student will get positive learning experience (15.M.).
R10 Application is easy to use (17.A.).
R11 It is safe to use (17.G.).
R12 Moving does not cause motion sickness (17.H.).

Based on the questionnaire and interviews, it was found fifteen statements, that teachers considered very important (M ≥ 4.5). These can be seen in Table 4. The reliability of the “very important” group was tested to ensure that all statements chosen were considered equally important among teachers. In general, the statements that are chosen for reliability analysis are within the same theme. In this case a slightly different approach was chosen. It was discovered that statements: 17.J., 15.G. and 15.F. had low item-total correlation (< 0.2) which refers to problems inside the group. After removing these statements, the reliability of the group was in a level that can be accepted (Cronbach’s alpha 0.72, n = 12). Finally, the “very important” group included three statements that related to Q13. Educational content and the teacher’s experience (25%), six to Q15. The student’s experience (50%) and three to Q17. Ease of use and technical solutions (25%) as listed in Table 5.

Table 6. The statements evaluated “important” in the teachers’ questionnaire (3.5 ≤ M < 4.5).

13.C. Material considers diverse learners.
13.D. Student gets feedback.
13.F. Students learn to use VR glasses.
13.G. The teacher can monitor students in the VR environment and see the progress.
13.J. The material supports investigative learning.
13.M. Student learns methods to collect biological data.
13.N. Student learns to apply biological knowledge.
13.O. The application provides a calm learning environment.
15.C. Student can use creativity
15.H. Students can collaborate, discuss and find solutions together
17.B. There is a tutorial that guides when using at the first time.
17.E. There can only be seen functions that are currently needed.
17.F. The environment resembles the real world.
17.I. The usage of the application doesn't expose students to game addiction.

From the group of statements that is categorized important ($3.5 \le M < 4.5$, n = 22) needed to remove eight statements because of low item-total correlation values (< 0.2). After removing statements 13.B., 13.K., 13.I., 13.P., 15.E., 17.C., 17.D., and 17.M., the reliability was accepted (Cronbach's alpha 0.86, n=14). The important group consisted mainly statements related to the educational content (n = 8) and the ease of use and technical solutions (n = 4). Only two statements related to student's experiences as listed in Table 6.

Table 7. The statements evaluated "undecided" in the teachers' questionnaire ($2.5 \le M < 3.5$).

13.H. The teacher is actively guiding students.
17.K. The application is used with touch controllers.

Three statements were found to include in the undecided group ($2.5 \le M < 3.5$) and after removing 17.L. (total-item correlation 0.13) the reliability of the group was accepted (Cronbach's alpha 0.63). Statements of the group related to teacher's experience (n = 1) and technical solutions (n = 1) as shown in Table 7.

Table 8. The statements evaluated "least preferred" in the teachers' questionnaire ($1.5 \le M < 2.5$).

15.I. Students can see each other in the virtual environment
15.J. Student can see the teacher in the virtual environment.

In Table 8 are presented the group of the least preferred features ($1.5 \le M < 2.5$) (somewhat disagree) which include two statements (Cronbach's alpha 0.96) from student's experience theme. Finally, to ensure that the groups formed had statistically significant difference from each other, it was calculated mean values for each group (i.e., very important, important, undecided, the least preferred). The groups were then compared to each other with MANOVA analysis after Mauchly's Test first confirmed sphericity ($p = 0.113$). $F (3, 51) = 41,33$, $p < 0.001$. It was recognized that there is a significant difference, and null hypothesis can be rejected as seen in Table 9 and Figure 15. In pairwise comparison was found a significant difference between all groups except 3 and 4. Because of similarities between groups undecided and somewhat disagree combining these two groups can be an option to consider. However, the group very important had statistically significant difference to all other groups and the statements of the group could be chosen as quality requirements (R1–R12 shown in Table 5) and features that will be implemented into final VR learning material and the application build. Based on these requirements found, the core need of VR learning material can be considered being to have a useful and safe application that is enjoyable to use.

Table 9. In MANOVA analysis are compared differences between category groups. Descriptive statistics and significance values of pairwise comparison (p =<0.001–0.036).

	N	Mean	Std. Deviation	p
Strongly agree	18	4.73	.255	**<.001**
Somewhat agree	18	4.03	.651	**≤.001**
Undecided	18	3.00	.624	**<.001**–.275
Somewhat disagree	18	2.36	1.068	**<.001**–.275

* The mean difference is significant.

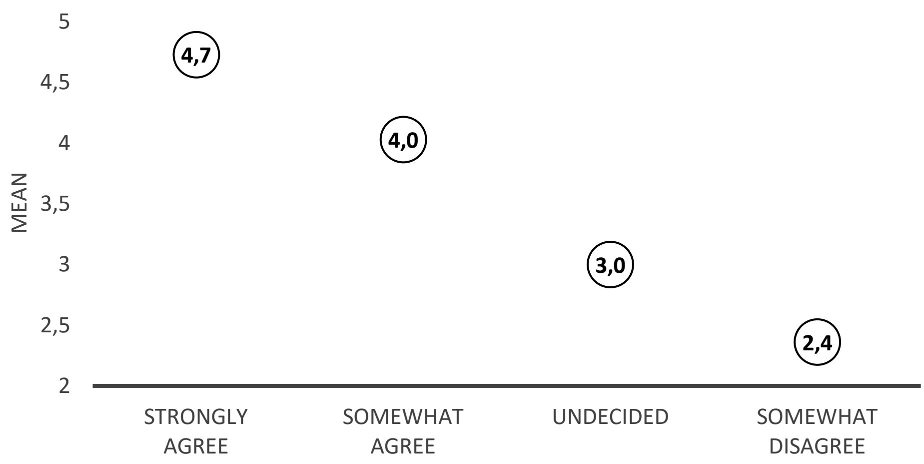


Figure 15. The mean values of four categories of teachers' preferences studied

4.2.3 Teacher's role, demographics and learning conditions

It was recognized that the group of teachers who participated in the study was heterogenous, but also was a good representative of the target population (i.e., biology teachers) of the study with different backgrounds which was the original aim. Teachers were undecided of the role they should have (RQ2) and passive role is preferred a bit more than active ($M = 3.29$). However, the standard deviation was a bit higher in this theme implying a variance of teachers' responses. The participants were asked demographic questions to discover, whether there can be found patterns related to, for example, preferences and age. The sample size of the collected data remained however quite small ($n = 18$) and to compare demographic relations, categories needed to be combined to make statistical comparison feasible. The aim was to form two groups with the same size (about nine participants per each). It was found from the demographic data shown in Figure 16 that the only measurement that enables this type of division was teaching experience. The first, group A, was formed from teachers that had teaching experience of a maximum of ten years ($n = 8$), and the second, group B, was formed from teachers that had experience more than ten years (n=10). With this division it was possible to explore the research question RQ 3: “Does teaching experience have an impact on preferred VR features?”.

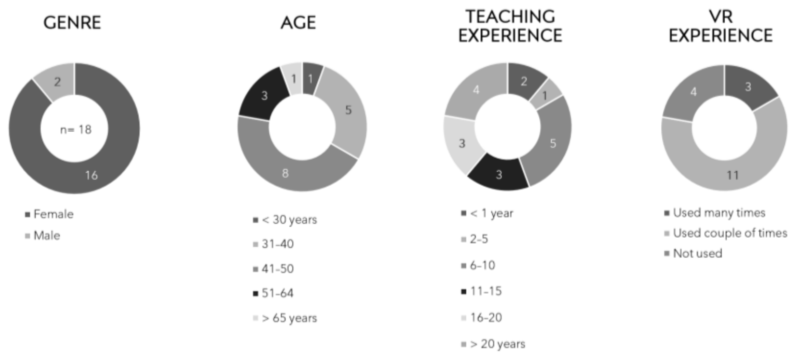


Figure 16. Demographic and background data of participating teachers (n=18).

Teacher T-test and Mann-Whitney U-test can be used to recognize whether there is a significant difference between the mean values of two groups. Teacher T-test is used with dependent datasets, for example if the attitudes of the same people are studied in two different points. Mann-Whitney U-test can be used with two independent samples with ranked scores, and the normal distribution is not a requirement. [228] As can be seen in Table 10, below, the data is not normally distributed but has mainly negative skewness (-1.05 – 0.00) and kurtosis (-1.99 – .13) values. Non-parametric Mann-Whitney U-test is useful for this kind of setting, and utilized in this study to make a comparison between the two group of teachers.

Table 10. Descriptive statistics related to themes and teaching experience. 1= Teaching experience 0 to 10 years (n = 8), 2 = Teaching experience over 10 years (n = 10)

	1		2		Mann-Whitney U	Wilcoxon W	Z	Asymp.Sig. (2-tailed)	r
	Mean	Median	Mean	Median					
Usefulness	3.99	4.03	4.47	4.50	13.0	49.0	-2.41	**.016***	-0.57
Usability and ease of use	3.64	3.64	4.16	4.21	23.5	59.5	-1.47	.14	-0.35
Safety and health	4.42	4.67	4.47	4.33	37.0	92.0	-.27	.78	-0.07
Teacher’s role	2.83	2.83	3.65	3.67	15.0	51.0	-2.25	**.024***	-0.53
Student’s experience	3.66	3.64	4.10	4.07	20.0	56.0	-1.79	.073	-0.42

*The significant mean difference

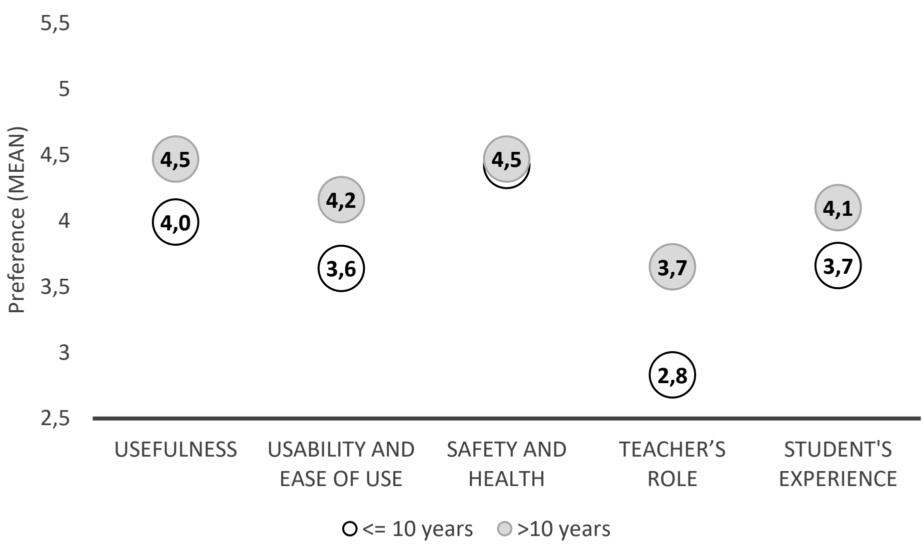


Figure 17. Comparison of mean values of Teachers having teaching experience over ten years or less.

Mann-Whitney U-test reveals that teaching experience relates to how teachers see elements important to include into VR learning materials presented in Table 10. The mean values measured from more experienced teachers are higher across the board as can be seen Figure 17. Teacher's with over ten years teaching experience evaluate the Usefulness (U = 13, n1 = 8, n2 = 10, p = 0.016, r = -0,57 (large negative effect size)) and elements that enable teacher's more passive role (U = 15, n1 = 8, n2 = 10, p = 0.024, r =-0,53 (large negative effect size)) significantly more important than teachers with less teaching experience. The significance of student's experience (U = 20, n1 = 8, n2 = 10, p = 0,073, r = -0,42 (moderate negative effect size)) is also very close to the usually applied 5 % level of significance. In this case the value is 7%. The r value 0.1 refers to small effect size, 0.3 is moderate and 0.5 large effect size. Plus, and minus signs relate to positive or negative correlation. [227] [58]

How to implement VR into education

Teachers were asked in open-ended questions of the survey about what kind of VR sessions they see beneficial and how sessions should be organized. The aim was to solve the research question RQ 5: "How can VR learning be implemented into education?". The topics for VR lessons that teachers preferred related to ecosystems and human anatomy as seen in Table 11. Responses that had the highest frequency related to ecosystems (n=11) and multiple topics that can be considered subtopics of ecosystems were mentioned. Species identification (n=9), vegetation zones (n=7), fieldwork (n=4), the carbon cycle, and human impact (n=2) were mentioned multiple times and increased the frequency of ecosystems to 33. Based on the results of this study the topic for the VR application is chosen to be a combination of three areas: the lake ecosystem, species identification, and fieldwork.

Table 11. Teachers' suggestions as a topic for a VR lesson.

Topic	n
Ecosystems (water/lake/ocean/bog), underwater life, interaction of organisms	11
Species identification, organisms e.g., plants, structure, cell	9
Vegetation zones, forest types, vegetation zones of the shore, forest layers, seasons	7
Fieldwork, excursions (e.g., forest), water sampling, the use of microscope	4
Carbon cycle and human impact	2
Human body, anatomy and organs	8
Evolution	1

The teachers were then asked about the preferred length of VR lessons, and how often they consider these lessons should be arranged. The answers are presented in Figure 18. The length of session that teachers preferred varied a lot, however the length of less than ten minutes was mentioned only ones and lengthier options received more agreements. Teachers considered that a suitable length for the VR lesson is from 11 minutes to the maximum of 40 minutes. 11–20 min (n = 6), 21–30 min (n = 5), and 31–40 min (n = 6).

17% of teachers reported, that they had already used VR headsets a couple of times with students in school, but others had not this kind of experience. The third thing teachers were asked about was the frequency of how often VR learning should be arranged. Most of the teachers considered that biology lessons that include VR technology could be arranged five to ten times (n = 8) or a couple of times (n = 8) during the school year as shown in Figure 18.

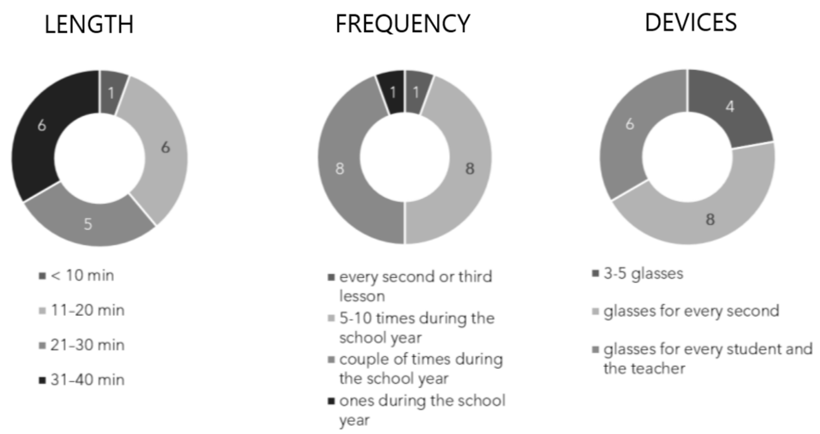


Figure 18. VR lesson. The suitable length is from 11 min to max 40 min. Preferred frequency is 3 to 10 times per school year and according to teachers, headsets should have at least for every other student.

To identify the environmental conditions of VR lessons, teachers were asked about devices needed. 78% of teachers preferred having VR headsets for every other or every student and the teacher. Only 22% of teachers considered 3 to 5 VR headsets to be enough in a classroom of 20 students. Teachers who selected the lower option reported, for example, being concerned about the price of devices and the possibly relatively short lifecycle before the technology becomes outdated. Not being able to install new applications or not to have software updates worries teachers. This was reported being happened with, for example, iPad devices that many schools have obtained. Trashing otherwise working devices is wasting limited resources and increases environmental impact. This was a concern that teachers addressed and it was the first challenge found (RQ6) in this study. Even though teachers reported their concerns, they also expressed interest in this relatively new technology and were hoping to have a useful and usable alternative for learning through VR.

5 Application – Learning material design

Systems designed should be comfortable and effective to use, effortless to learn and easy to remember [208]. Good application is learnable, memorable, efficient, risk minimizable, and agreeable [210]. It is attractive, comprehensive, takes cultural norms into consideration and meets people's expectations [207]. People favor using software that is more usable [29]. Janet Murray has said "Users do not think in terms of abstract affordances, but they expect new digital artefacts to have features similar to other digital artefacts" [160]. To understand the needs of users and fulfill the educational purpose is important [207]. With educational applications, materials must be complementary and mutually reinforcing the educational program and testing before use is important. [207] Making a usability evaluation in an early state can be cost-effective and reduce the need for bigger and more expensive fixes later [231]. It is seen important the personnel of the field (e.g., education) to take part in the design process and when the product is ready for delivery, sufficient instructions must also be included to ensure effortless implementation. [207] To deliver a useful IT artifact, defining requirements is essential. Requirements specify the needs of the user and relate to solving problems and meeting conditions or capabilities [14].

In the Chapter 5, the focus is on the application that is designed for testing purposes based on the preferences of teachers. At first the device used and development tools utilized in the development process are presented, after that the requirements and the implementation methods and means are detailed.

5.1 Equipment and development tools

Multiple VR devices are today available with various qualities. In this study, the Meta (Oculus) Quest 2 headset is used. The model can be used without a PC connection but has an option to deliver the experience to other participants through smartphone or on computer's display with Oculus app. However, this feature was not used in testing. The headset has a compatible price but even sufficient qualities. If preferred the headset can also be connected to a PC with wire and utilize applications that require more capacity. Self-made applications can also be transferred to the device via cable connection. Meta Quest 2 headset is a 6DoF (degrees of freedom) device. The number of freedoms refers to directions of possible head movements tracked by the device as visualized in Figure 19. The smartphone requiring VR devices, for example, Google cardboard, which were popular in the beginning of 2000s, had

3DoF (pitch, jaw and roll). The tracking is more precise in 6DoF headsets, and the experience is more immersive.

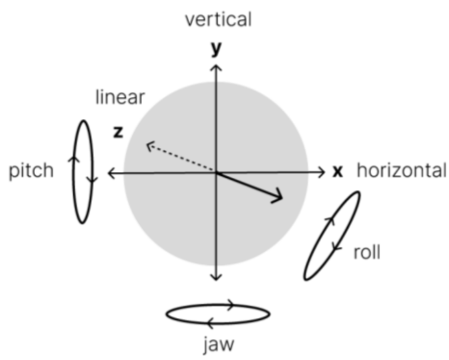


Figure 19. 6DoF include horizontal, vertical and linear movement as well as pitch, jaw and roll.

Two hand controllers are to interact and to control the VR experience, but the Quest 2 headset can also be controlled by hand gestures. In this study hand controllers are used since this was preferred a slightly more in the questionnaire by teachers. The Quest 2 tracks head and body (e.g., hands) movements, and all needed sensors are internal. The headset has an LCD (Liquid-Crystal Display) display with resolution of 1832 × 1920 pixels per eye. The Quest 2 can also be used with glasses on if an add-on piece that is included in the normal set is used which is beneficial and enables to meet the different needs of students. The headset has 3D positional audio and can also be used with external headphones if preferred. The ready-made content (applications) can be downloaded from the Meta Quest Store or unofficial stores such as Sidequest VR [232]. Application can also be self-built by, for example, Unity engine by enabling headset's developer mode. There are also platforms that allow easy build such as Meta Horizon worlds [33]. Developing applications from the start requires the use of the Meta account, developer mode and to have a computer powerful enough to run VR applications. Developer mode is utilized to develop the application used in the testing of this study. The headset's requirements for the computer can be confirmed from the manufacturer or seller. [162]

Development tools

Two engines that are usually utilized to develop 3D VR games exist: Unity and Unreal engine. Unreal is often used if the aim is to develop games that have the best graphics and realism. Unity's graphics are not that advanced, but otherwise it has alike features. Prizing policy is quite similar, and both are free to use until certain limit (Unity: gross revenue of

$100,000 and Unreal $1M). [233] [234] Because the aim is to build an VR application, extra-quality graphics are not something that is required in this case. Often games which use these high-quality graphics requires also a powerful computer that the headset is connected to. Now, the application is designed for a standalone headset (Meta Quest 2) and Unity offers required graphics tools for that purpose. Other reason that favors the selection of Unity is related to its familiarity. Unity is chosen for the engine to be used in the development process of this study.

5.2 The implementation of requirements – methods and means

The study utilizes the participatory design process, and the learning application is custom-made based on the preferences of teachers to meet this specific testing purpose. Requirements for the application are defined by the teachers' questionnaire and relates to five themes studied (i.e., usefulness (Uf), student's experience (E), usability and ease of use (Ub), teacher's role (T) and safety and health (S)) and each requirement can be combined with multiple themes. It is important the material to have a clear educational purpose and many of the teachers' expectations relate to student's experience as well. The theme usefulness, in other words the benefit aspect, forms combinations with other themes in most cases. Themes and theme combinations emerged from teachers' questionnaire are EUf (n = 5), E (n = 1), Uf (n = 2), UfUbT (n = 1), UbT (n = 1), SUf (n = 1) and S (n = 1).

Principles of IxD as well as elements of gamification and XR design are utilized in the design of an educational VR application and to implement requirements. How each quality requirement (R1–R12) is implemented into the application are introduced in details next.

R1. Material obeys the curriculum (Uf). The species identification is selected as a topic for the application designed. To be more precise, the material is designed to complement the exploration of the lake environment topic of lower secondary school students (age 14). Field explorations are part of the biology curriculum (R3) as well as species identification [93]. It is not always possible in schools to execute excursions in natural habitats for different reasons. This application aims to simulate fieldwork and to offer an alternative for schools, in case needed. Even though, it is important to remember that simulation cannot fully replace the experience received from the field.

The learning task of the designed application, is to find as many species as possible from the environment. The environment is very simple and includes a lake, trees, stones, vegetation,

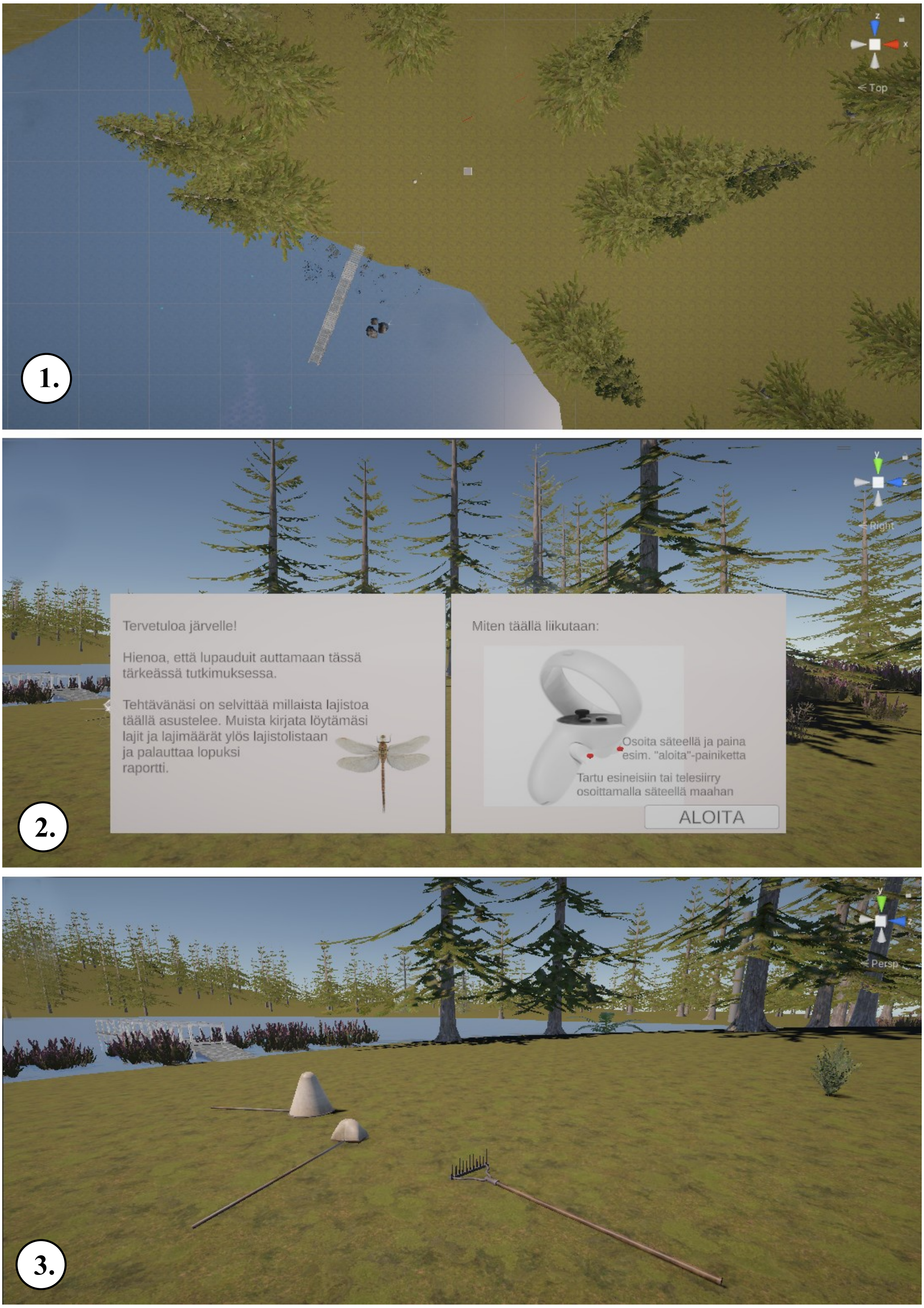


Figure 20. The VR learning environment of lake exploration. A bird's-eye view of the lake environment (1). Instructions about how to proceed in the environment and use hand controllers (2). Equipment available to execute the task of field exploration (3).

hills (land forms), a dock, info boards, info labels, an aerial net (butterfly net), a dip net (d-net), a rake, a species list table, animated spheres (resembling insects or mollusk) and species' information cards with rotating images of species as shown in Figure 20 and Figure 21.

R2. Help is provided when needed (UfUbT). In XR the guidance can be made through motion, audio that is spatialized, or by using colors. If a deeper impact is needed, the object can surround the user. It is recommended to offer an effortless way out of immersive experience with, for example, clear symbols. [235] It is a challenge to make people act in preferred way in a digital environment and scaffolding is needed. The need is similar as in a physical environment; however, digital input/output (IO) devices need extra attention. In the best scenario the guidance is not realized but is considered as a part of a game or material designed.

The guidance of a game can be done through environmental manipulation (e.g., only one route available), goal injection (e.g., someone is guiding to go to a place), shifting personality (e.g., love pets), plot points (e.g., when time is right the event happens regardless of user's actions) [236] or dropping the fourth wall (e.g., the user is instructed straightforward to do something) [237]. Dropping the fourth wall is the most direct way to guide. It can easily break the feeling of immersion and interrupt the experience.

More of the simulation style than gamification aspect is used in this application designed and scaffolding resample what is convenient for fieldwork in general. Two info boards to guide students' journey are placed into the VR learning environment as illustrated in Figure 20:2. Boards resemble interactive whiteboards which students in most cases already are familiar with, and the guiding is very straightforward. The info-boards are placed in front of the start point of the player in the VR application. The first info-board is for a short introduction and delivering the task students are asked to perform. The second info-board includes instructions on how to use hand controllers (e.g., what happens from each button). Both boards disappear, and exploration can begin by clicking the start ("aloita") button from the second info-board. The board can be set visible again by clicking a button with text "show instructions" ("näytä ohje") on a species list table (see Figure 21:9) that is lying on the ground nearby the dock. Species to be caught (blue squares) are moving up and down to draw students' attention. The outer line of square (e.g., insect) has a see-through feature so that species cannot hide behind, for example, trees and are easier for students to discover. Scaffolding is added to guide how to catch species and if, for example, dragonfly is tried to be caught with wrong equipment (e.g., dip net) an info label “change an equipment” (“Kokeile muuta pyydystysvälinettä”) will appear to guide the user as shown in Figure 21:6.

A species card will open in the middle of the user’s field of view, when the right equipment is used in catching. The place is chosen to tackle the challenges of VR and ensure the student sees the card [178]. Student can add the species into the species list table by clicking the orange arrow (download) symbol, after learning the species by clicking buttons (green plus-symbols) and opening the information texts. The colors have been chosen to draw students’ attention. An instruction: “First, find out what the species is” will appear if the arrow symbol is clicked before it is discovered what the species was. A text “added to the species list table” will appear, after insertion is succeeded, to inform the student. A text “did you remember to add the species to the species list table?” will be shown, if the species card is closed (clicked cross symbol in the upper right corner (standard layout)) without adding the species into a species list table so that the student will not forget the task the next time.

R3. Material intensifies the knowledge of students (Uf). The species selected for the learning material are the same as found in the study book of the class to be tested [238]. The learning material intensifies the theme earlier studied and teaches students to find and catch species with the right equipment in their natural habitat. For example, a dragonfly’s larva can be found in the bottom of the lake and adult dragonflies are flying in the sky. A species card with a rotating image of the species appears (see Figure 21:7), after catching a species (i.e., touching a blue square with a tool as illustrated in Figure 21:5). A question on the card appears asking “what the species is?” (“Mikä laji?”), and by clicking buttons on the card, the student can solve the name of the species. On the card can also be found an extra information about the species which cannot be found in the study book. Behind green plus-signs can be found, for example, how to identify species, and what are the special features of them as can be seen in Figure 21:8.

R4. Students experience things instead of standing by and watching (EUf). VR provides an immersive experience where students can interact with the simulated VR environment. Every student is using a VR headset, and they have a unique experience in the VR lake environment. To increase equality among students is one of the goals of technology usage in schools [93] and students can be self-experiencing and actively participating in the VR environment, which is beneficial in the terms of learning.

R5. Student can explore the environment (EUf). The environment designed in the study is relatively simple. The storyworld, where game events take place, can be linear (i.e., parallel with only one path to follow), branching (i.e., threaded storyworld where events have parallel paths to choose from) or open (i.e., scattered storyworld with no predetermined paths and user

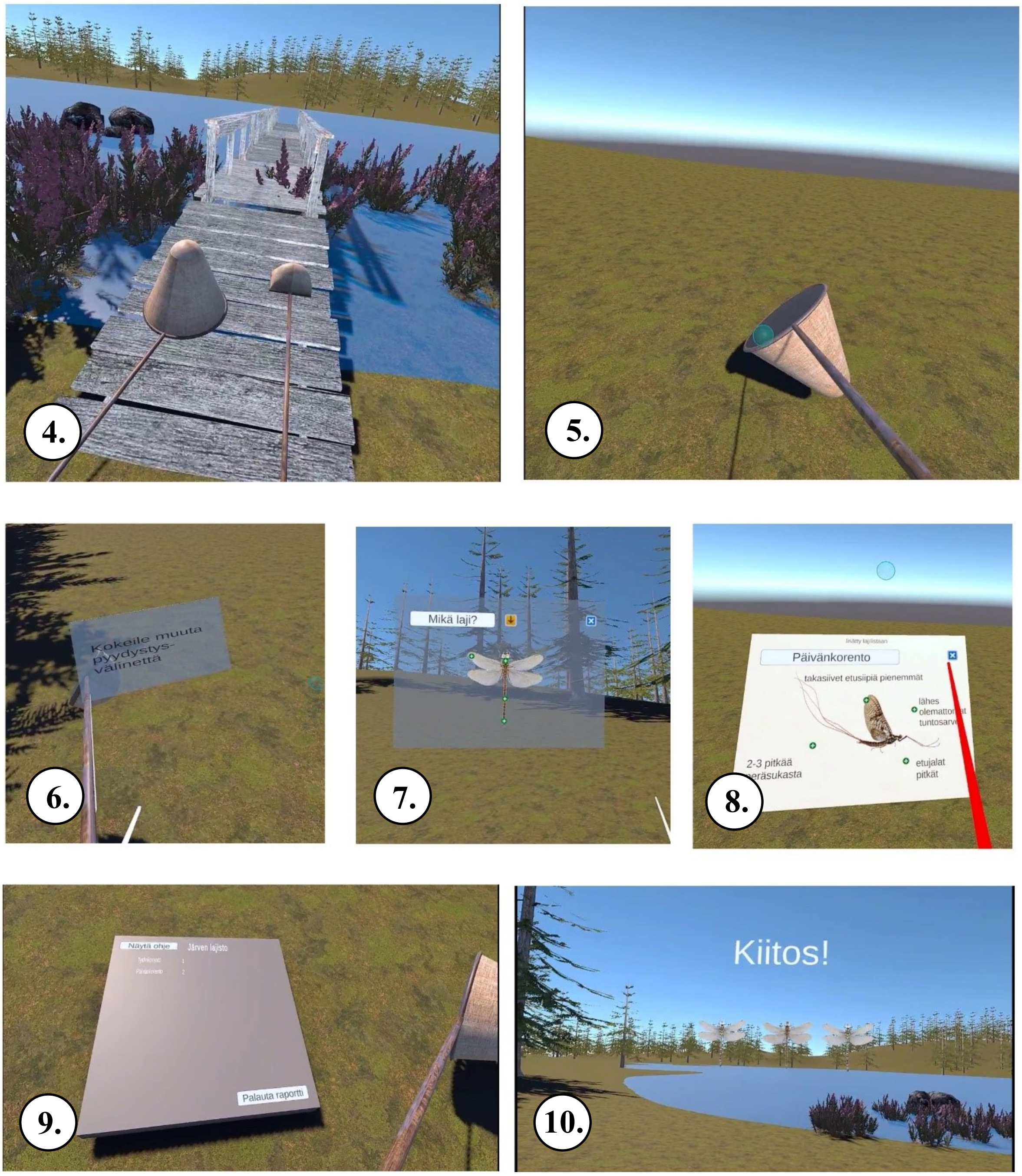


Figure 21. Catching equipment (an aerial net and a dip net) on a dock (4). An insect (blue square) is caught with an aerial net (5). A label instructs to change catching equipment (6). The species card (7). The species card with information being open (8). The species list table ("Järven lajisto") with buttons "Show instructions" and "Return report" (9). End and thank you ("Kiitos!") screen appears after returning the report (10).

can choose it) as illustrated in Figure 22. Threaded storyworld has one storyline with offshoots, in scattered the pieces of the story can be experienced in different order (the order is defined by the author or a character) and in converging storyworld every player have a role and impact on the final story. [239] [236] [178]

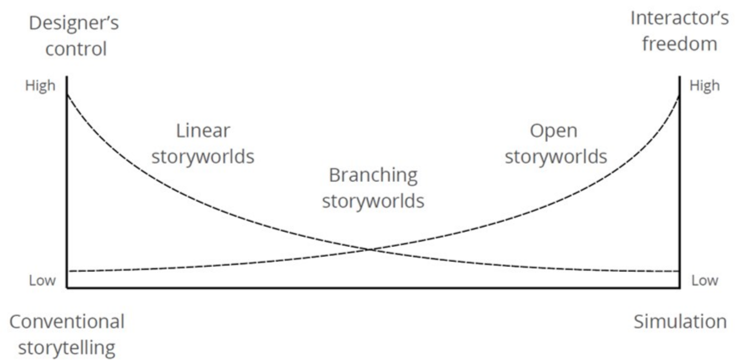


Figure 22. Different storyworld types by Smed et al. [178].

Different approaches are available regarding the control of game experience (i.e., author centric and character centric). The more possibilities the user has to choose different paths the more control and more character centric the storyworld is. The character centric (i.e., user centric) approach has gained more attention after the 2000s [240] [241], but how to solve the narrative paradox; give the user the agency and the freedom of choice, but same time have predetermined conditions and author control [242] and avoid a combinatorial explosion with too complex structure [236] [243] [244] [178].

In this study, students are able to explore the environment and move freely by teleporting. They keep collecting species until the teacher advices to return reports. Material designed is suitable for VR sessions with different lengths which is a feature, teachers reported preferring. The storyworld is open [239] since students do not need to follow specific order or path to proceed. The maximum of five species are visible at the same time and after one is caught another randomly selected (n=18) will appear in a random place (n=81). Students can progress at their own pace and if they prefer, they can even sit down and rest, for example, on the dock.

R6. Student gains new experiences (EUf). VR is quite new technology in schools and not many students have tried the immersive VR headset before. However, many possibly have an experience on a smartphone requiring VR devices, and this was at some level familiar technology. VR headsets are devices that can offer students new experiences. Students can visit, for example, museums, new cities or the International Space Station and experience things, which in other cases, would be impossible.

In this study, the focus is to familiarize students in VR technology and possibilities related to VR and fieldwork. The experience can be made more authentic by utilizing features of devices, and to find key moments to be emphasized is ideal. Certain features of VR can make the experience feel more immersive, and make objects feel more alive. The use of lights, shadows, and subtle depth can make the environment to be more real and interesting. By using animations and adding sounds for the atmosphere, a more immersive experience can be created. [235] In this study, animations are added to vegetation, and for example, plants are slowly swinging in the wind. Species to be caught are animated as well.

R7. The motivation to learn the subject increases (EUf). New technologies have been found to increase motivation especially in the beginning [120]. Adding gamifying elements in the environment is something that can enhance motivation as well. Non-linear and stratified structure, game levels, story, clear goal, suitable challenges, speed, competition, and immediate feedback are features of a successful game. [54] Games do not need to include a story, but adding one can improve the experience [245]. Narratives and stories are influenced by experiencer's individual and collective imagination. The story is different for everyone, which makes every experience unique [245]. It has been observed that adding printed text to narration does not significantly have an effect on learning. However, text in forms of narrative is less overloading than text only option. [176] Makransky et al. (2019) found that by using immersive VR can be improved perceptions, for example, science labs, but resulted in minor learning since students receive cognitive overload. Another issue that can cause cognitive overload in the VR environment is that not enough time is allocated to process new information. [176] Students see a short introduction story in the first info-board:

"Welcome to the lake! Great that you volunteered to help with this important investigation. Your task is to discover what kind of species live here. Remember to write down on the table of species all the species you find, and species amounts. In the end, remember to also return the report."

The introduction story aims not only to deliver the task, but to get students more interested and motivated to complete the task. The learning material follows the inquiry learning approach [119] [120] and students are not informed straightforward how to find and catch species (i.e., to choose and take an equipment and catch blue squares). This is considered as the challenge students need to figure out. The predetermined goal is in this case to find all species of the lake environment with equipment available. Students are encouraged to investigate and try different things. They receive immediate feedback whether the action was correct (e.g., species collection) or not. Collecting species can be considered as points which is a gamifying element and students can follow their progress on the species' table. They do not see each other nor the progress of others since this was among one of the teachers' least preferred features. However, they can compete with each other if wanted by communicating and discussing the results.

R8. Student can experience something that is not possible in real life e.g., fly (E). VR can offer new experiences and abnormal qualities because real life restrictions do not apply in the VR environment. Teleportation is something that is not possible to do in real life and in the application, students move by teleporting. Another abnormality is the possibility to walk through the lake. The vegetation of the VR environment was designed also to be a bit bizarre with giant heather plants that were growing not only on the ground but water as well. The size of heathers enabled students to see the plant structure better and with more details. The abnormalities were designed in the environment to increase curiosity and interest.

R9. Student will get positive learning experience (EUf). VR systems can increase the motivation of students and help to concentrate on the subject learned by preventing outer distractions [246]. VR technology can be utilized for learning concepts, definitions and theorems [246] and appears to be a useful tool to use for weaker learners and learners with cognitive, social or physical impairments [246] [11]. For example, students with cerebral palsy can experience walking for the first time or VR technology can provide visible sound solutions for people with hearing disabilities [11].

Flexible design is important in XR since surroundings fluctuate and using smooth transitions is beneficial. Adding soft edges in digital content helps objects to blend in the background. [235] which was a feature used in this application as well. In the application designed theory of the study book was connected to fieldwork simulation. Practical experience and doing concrete things are often something favored by the students. However, using new technical devices can stress students [118] and making the environment easy to use but still challenging was the aim and expected to result a positive learning experience. The needs of diverse learners are aimed to meet with a simple design and by giving every student an undisturbed digital space to concentrate on the task given.

R10. Application is easy to use (UbT). Navigation and finding objects in a new application should be as easy as with interfaces people already are familiar with. “Good design is always human-centered”. The most important content should be placed at the center of the user’s view and follow the angle of user’s head (ergonomics). The content should be far enough so that users cannot reach it with hands. Objects that are placed near invite people to interact with and placing controls near the object is recommended. Giving the user the option to resize and replace objects is found to increase the comfort level. With design, it is good to remember that less is more and not fill environments with objects. People’s focus can be guided as well to reduce feelings of overwhelmed, and for example, dimming the background of XR applications helps to concentrate on the preferred content and ease the use (i.e., operability [229]). [235] The design of the VR environment of this application is kept simple to avoid the feelings of overloading and overwhelming. The content (e.g., info-boards and species cards) is placed in the middle of the users’ visual field and set at suitable distance to enable effortless usage.

Hand controllers of VR headset are equipped with buttons similar as game joysticks and starting the usage is possibly easier for students interested in gaming. Only two buttons of hand controllers are active and can be used for proceeding in this application. A trigger button enables to press buttons on boards and tables. A grip button enables grabbing an object or teleportation. The instructions how to use hand controllers and what to do in the VR environment are very short to read to ease the usage.

R11. It is safe to use (SUf). The safety aspect is something that is especially important to consider in schools. Particular features of VR can cause safety issues and are important to consider. Physical safety is, of course, important. Because normal sight is blocked by the VR device, it is important that the headset has qualities that enable safe usage. Colliding with other people, furniture, walls or falling stairs needs to be prevented. Meta Quest 2, the headset chosen for this study has a guardian system (i.e., hazard warning [229]) designed for this purpose [188]. If the user is near exiting the play area, a digital wire fence will appear in the visual field and in case leaving the area a see-through feature activates and the real environment as black and white can be seen through the device’s cameras. Two options of the guardian system exist in Meta Quest 2. The play area can be drawn into the room or used a stationary boundary. A stationary boundary (about square meter play area) can be chosen, if there is very little space to move around or students need to be sitting or standing in place when playing. The selection can be done from headset settings. Physical indicators such as mats can be used on top of the options that the device offers. The player is no longer inside the play area, if the mat cannot be felt under feet. Regular mats or special mats made for VR gaming purposes can be used. As a game design aspect designing games with requirement of minimum movements and fast actions can improve physical safety, however, interactivity is essential feature of VR and following events (similarly as watching TV) through VR headset can be considered as not fully utilizing the qualities of the device.

Another issue related to VR is that eye fatigue is evidential even after 10 minutes use of headset [247] Offering possibility to have breaks while using headset can help this problem or designing material for only short VR sessions is an option as well. The application designed for this study can be used as short periods as required and there is no predetermined length for the session. Having breaks while playing does not interrupt the progress of the game neither.

Third safety and health issue, that has been discovered related to VR gaming, is that it induces depersonalization-derealization disorder (DPDR) -like symptoms. significantly more often than regular PC gaming when measured immediately after play session. DP refers to alienating and unrealness feelings of self. DR refers to the detachment from reality [248]. However, after one day or a week from gaming the difference has disappeared and no long-term differences between VR and PC gaming have been found [248]. The use of colors must also be considered carefully. It can attract users and draw their attention, but also produce emotions, for example, pleasure, tension, fear, and anger. [160] In schools, VR sessions are most often quite short, but with longer sessions, some players may suffer this kind of symptoms shortly after using headset and is something to consider. In this study, it is avoided to design emotionally loaded events and material designed is selected to be very neutral and real life resembling to make students feel secure and comfortable in the environment.

R12. Moving does not cause motion sickness (S). Cybersickness can be experienced in VR and is caused by the conflict of the stimulus of the vestibular system in the inner ear and the visual system [249] For example, flying around in a VR environment is something that can cause cybersickness. Design choices have an effect on the experience. In the study, it is paid special attention to avoid feelings of cybersickness. Free movement was disabled from hand controllers, and teleportation was the only option to move unless physically moving, since it has been observed that free movement in average, can cause more motion sickness for people [250]. For the same reason, turning around is enabled only by physically turning a head and moving.

6 User Testing

User testing includes meaningful activities (i.e., tasks) that participants are conducting while being observed [251]. The testing can be done in different environments and under varying conditions for different purposes, for example, to verify design solutions, to eliminate design problems, to improve profitability, to discover opportunities, or to learn about users [251] [97] [252]. User experience (UX) is in key position in product design, and it relates to usability, usability testing and user-centered design (UCD) [97]. Morville's *honeycomb* model includes seven facets of user experience: usefulness, desirability, accessibility, credibility, findability, usability, and value. The model can be used to design qualities of experience. [253] [254]

Utilizing user tests is a way to discover the needs of users, provide more comprehensive user experience, and to get external opinion, and a backchannel before delivering a product for users [160]. Testing includes preparing a test plan which contains the scope, approach that is aimed to utilize, timetable and detailed information about what will be tested, how and by whom and whether there are risks related to testing session that needs to be considered [14]. The four approaches that can be used in the user testing are exploratory (i.e., formative), comparative, assessment (i.e., summative) and validation (i.e., verification) approaches. This study utilizes the exploratory approach which can be used in an early stage of design process and found vital to solve, for example, users' thoughts related to the product. [97]

In this study, the aim of the user testing is to discover what students think about the VR learning material designed, what is the user experience like, what kind of features users feel important, is the application easy to use, is it useful and how the implementation of teachers' expectations has succeeded. It is later compared the extent of which the features preferred by students are similar to teachers' expectations. Before the testing can be performed, a research permit from the school, which students are to be tested needs to be applied. The permit includes detailed clarification on how the testing will be executed. Another very important preparation needed to perform is to obtain the testing equipment and to formulate the feedback questionnaire (see Figure 23).

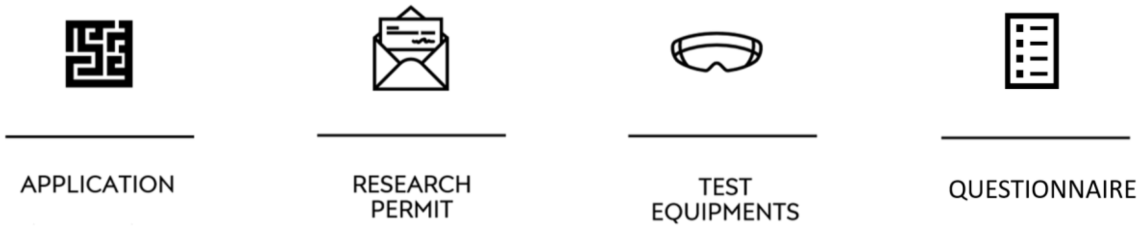


Figure 23. The preparations and components needed for the user testing.

6.1 Methods and data collection from students

In this part the methods which can be used to study students' learning and the usability and usefulness of the learning materials are discussed. It introduces the methods and data collection of this study related to students' learning.

6.1.1 Methods

Various methods to collect data have been used in previous studies related to learning. Questionnaires, observations, and students' diaries have been used to collect qualitative data on learning [20]. In *game testing* can be used testing disciplines such as *usability testing*, *playtesting* (i.e., fun testing), *balance testing* (i.e., are game conditions fare for both human and AI player), *localization testing* (i.e., to fit the game into another geographic region requires e.g., translations), *compatibility testing* (e.g., what devices, software and web browser can be used) and *compliance testing* (i.e., certification is required before becoming in the market (hardware developers e.g., Microsoft conducts this phase)). Testing techniques that can be used are *Ad-hoc testing* (i.e., free-form testing), *progression testing* (i.e., playing from the beginning to the end), *automated testing* (i.e., computer testing), *functionality testing* (i.e., game functions and looks as designed), and *regression testing* (i.e., making sure old bugs have not returned). [255] Usability testing and Ad-hoc testing are utilized as testing disciplines and techniques of this study. In the user testing, however, students are given a task to deliver, but the instructions are not very strict.

The Research frame can include one or more separate groups to be tested. All et al. (2016) have considered that in *digital game-based learning* (DGBL) the need for a control group (i.e., no manipulation) is not a compulsory element in testing, especially when no comparison with another method is involved. Three separate groups are suggested if the aim is to examine, for example, motivation instead of purely cognitive aspects. Three groups used can be, for example, group 1: game with learning content, group 2: traditional learning practice, and group 3: game with no learning content. Pre-tests are seen as important for discovering differences between control and experimental groups, tracking down the learning gains, and finding out characteristics of dropouts. For example, dropouts are not a random group but often people who have difficulties conducting tasks. It has been noticed that the results of learning should be evident not only in post-tests (cf. follow-up study) but also in school grades. This is more likely to be seen in longer-term interventions and can be found by using pre- and post-tests. The idea of pre- and post-tests is not that people are taking the same test twice since they normally perform better the second time (i.e., the practice effect), but to have two similar tests. Using similar pre- and post-tests is recommended since comparing results is easier if, for example, the type of questions and difficulty level are alike. [56] Testing the similarity of the tests beforehand is a choice to consider and confirm the correspondence of scores and themes. [56]

User testing can be carried out with different settings, for example, in a laboratory, remotely at home, or by performing field tests. In many cases, controlled laboratory setup is preferred instead of, for example, remote gameplay at home since tracking and making observations is more difficult that way. In a home, conditions of the testing environment can vary and have an impact on the results of participants. Laboratory testing has been discovered, especially useful in the first phase of the study. In this study, the user testing was decided to be performed in an authentic classroom context, since field-testing is a method to utilize, if deep understanding of users in their normal environment is needed. In the field, it is possible to study user’s real behavior in the context related. Exploratory field-testing produces more data since environmental data needs to be considered as well. Even though it is in many cases advantageous to study users in their natural environment, it can rise challenges such as disruptions that may occur and cannot be prepared or controlled. [251]

Traditional the user testing requires accuracy and lots of time and money. The number of participants has typically been 30 to 50. [97] However, Nielsen & Landauer (1993) have discovered, after making use of the Poisson model, that 3 to 5 participants produce the optimal cost-benefit ration and most of the usability problems can be found without the need to involve more people. The conclusion is based on multiple separate interface debugging tests utilizing a heuristic evaluation or a user testing (i.e., subject related tests) as methods. The studied systems were diverse, and in some systems the usability problems were easier to find. Severe problems discovered the easiest in both the user testing and in the heuristic evaluation [256]. In heuristic evaluation, the proportion of found usability problems increased when professional evaluators were used. Three regular usability specialists found 75%, and three double-specialists over 90% of usability problems. However, it was required 14 novice evaluators to meet 75% volume and with five novice evaluators half of the problems were found. [257] In addition, the results of subject-related user tests are lower than heuristic evaluations implying the need for a bigger sample size. The bigger sample size is acquired as well if novice users are tested, which in many cases is intentional and appropriate [258]. In this study students who are intended users were decided to be used as testers of the application as well.

Using randomization or matching are preferred methods in studies [56]. Convenience sampling and purposive sampling [58] is utilized for selecting participants for the user testing of this study. It is considered that one class of students is an appropriate sample size for the study [258] and most usability problems can be found and student's opinion surveyed. The class to be tested is selected since having similar interest in new technologies (i.e., study in a science-oriented line), as the teachers who voluntarily took part in the survey probably had. The fitting class for testing purposes was found in the school which is a multicultural and international school and has a science-oriented line. The school is a part of research, experiment and development (TutKoKe) networks of university training schools [259].

In research permit that was obtained from the school clarified, for example, for what purpose the study is utilized (i.e., in this case for master thesis), who is applier, what is research plan, how old students aimed to be tested are, what is the time frame, impacts and other resources required from the school, who is supervisor and whether the commitment to follow the principles of research etiquette is accepted. The research permit was received for the spring semester of 2023 for a class (n = 17) of seventh grade students (age 14) who studied in a science-oriented line.

Next it was needed to acquire the testing equipment, and two options existed. It appeared to have a possibility to lend ten Meta Quest 1 headsets from another university. Because the class to be tested had 17 students, this was an ideal option even though the model was an older model. Since teachers had reported in the questionnaire, preferring to have VR headsets for every or every other student, the research permit was applied to test half of the class at a time during two lessons. However, lending gear from another university appeared to be impossible because of the insurance policy of the Department of Computing and testing needed to be executed with four headsets available. This was unfortunate, shortened the testing time per student, and impacted the usefulness of the study. It has been discovered in previous studies that chosen equipment can have a significant effect on results. With sampling, it is recommended to extend the period of testing instead of performing short tests when considering behavioral research after it provides more useful data and accuracy on results. [58]

The testing was performed with four Meta Quest 2 -headsets and the class to be tested was divided into five smaller groups of 3 to 4 students. The group dynamic was considered in grouping and was conducted by the teacher of the class. The same grouping was used as in previous lessons, and students were familiar with working in these groups. This increased the authentic classroom context experience during the testing. Because of the sample size (n=17) and a short testing time for the focus of this user testing became more the usability, satisfaction and experience aspect rather than learning. There was no need for having a control group, and all students were similarly tested. In behavioral science *case study* (i.e., naturalistic observation study) is one observational method where individuals or events are studied [58]. This study utilized a case study setup.

6.1.2 Data collection

According to Nielsen, the core elements of usability testing are the facilitator, tasks, and participants. The facilitator administers and delivers tasks for the participant, observes the testing situation, collects feedback, and asks specifying questions in case needed [252]. The participant observation provides valuable information inside the research setting, for example, friendships and other issues that can have an impact on results. Recorded audio and video are recommended to support data collection with observational methods after enabling revisits and checks made later. [58] Using standardized tests is a practice suggested. Otherwise, clear descriptions of how the test is constructed and carried through are expected. Four elements needed to be explained are (1) a type of knowledge measured (e.g., ability of problem-solving), (2) a type of questions and (3) the difficulty level and (4) how the scales used relate to the psychometric properties (i.e., individual parameters e.g., personality [260]). Metrics of usability are efficiency (i.e., systems capability to perform functions designed [14]), effectiveness (i.e., accuracy, speed and completeness to reach goals [251] [15]) and satisfaction [15] [97]. The five measurable human-factor goals can be used to examine success in predetermined benchmark tasks. The goals are (1) time to learn, (2) the speed of performance, (3) the rate of errors made by users, (4) retention over time, and (5) subjective satisfaction. [160] Regarding game testing and observational methods, it can be calculated, for example, how often students expressed happiness, sadness, fear, anger, disgust (i.e., fundamental emotions) during testing [261] [262]. Related to system usage, continuous data can be collected about the quantity of needed aid, and the features used for the most often as well [58] [160]. In this study, because of the environmental factors and field-study-setup, it was decided to focus on making overall observation during the testing process and to collect feedback from participants with questionnaire afterwards.

Figure 24. Phases of students' user testing

The testing was conducted on Friday 28th April in 2023. The user testing consisted of three parts: instructing, testing, and getting feedback as shown in Figure 24. Time for testing was 2 x 75 min (a double lesson). There was a break between lessons, and the second lesson was divided into two sessions (55 min and 20 min sessions) because of the lunch break. Five student groups were formed, two groups of four students and three groups of three students (n=17). It was allocated 20 to 25 min time for testing of each group and in the lessons was employed the station rotation model where groups of students rotate through activities and follows predefined schedule.

Adding an introduction to testing session is a recommended practice and it is a tool to decrease guidance need. If debriefing is preferred, it is an option after the post-test since it does not impact the results. Each participant should get similar conditions and information during the testing session (i.e., intervention): time, content, support, environment, awareness of testing and a reward from participating. [56] At the beginning of the first double lesson and before testing sessions, it was arranged a short introductory for the whole class where everyone introduced themselves, and the class was informed about the context related to the testing as well as other formalities of the lesson. The testing situation was made as comfortable as possible and embedded inside normal classroom work. Testing was conducted at the school and was connected to topics previously learnt. The VR testing was among four other tasks that each group of students was to deliver during lessons. The teacher of the class was responsible for guiding other activities and sent a new group for VR testing when the previous finished.

Before testing, each group received more detailed instructions and information about the testing. They were informed that the activity of the station was to test a VR learning game. It was allowed to discuss with each other and students were encouraged to act as in a normal class situation. In addition to written instructions found in the VR environment students were encouraged to discuss with each other to solve appearing problems and to ask for help from facilitator if needed. However, first, the group should try to solve problems without external advice. Students were informed that they may experience nausea while using VR headset and told that at any point if chosen they can take off the headset and have a break or even terminate the test if preferred. After that, it was demonstrated how to wear the VR headset, use hand controllers, and informed about the play area and safety features of glasses. There was a guardian in the headset as well as a real carpet marking the boarders of the play area. The carpet revealed to be useful since it appeared during the testing process that two of the headsets had the guardian disabled and needed to be manually enable in later testing. Students were informed to have about 20 minutes to go for testing and after that they are asked to give feedback by completing a questionnaire. Students were asked not to discuss with others students in the classroom about the details of testing after they return to enable classmates to have similar experience as they had. Then it was asked whether students had some questions about the testing. The facilitator then started applications of all testers and helped to wear the devices.

The testing was carried out in the corridor just outside the classroom during the first lesson and followed the test setup A illustrated in Figure 25. Because of that, it was not possible to start the preparations and building test setup required before all other lessons had started and the corridor was empty of students. The preparation and time required for the introductory part shortened the testing time available in the first lesson. In addition, the number of disturbances caused the passer-by increased towards the end of the lesson and continuing testing in the next lesson was practical. Finally, it was tested two groups in the first lesson. The classroom was changed between the first and the second lessons, and the testing place changed as well. For the second lesson, the teacher of the class had been able to book a room next to where the testing could continue. In the second lesson, two groups were tested before the lunch break and one after. The testing time for the last group was unfortunately short.

Luckily, the group was able to use a part of their break time to fill in the feedback form. Each group spent 20–30+ minutes testing and giving feedback.

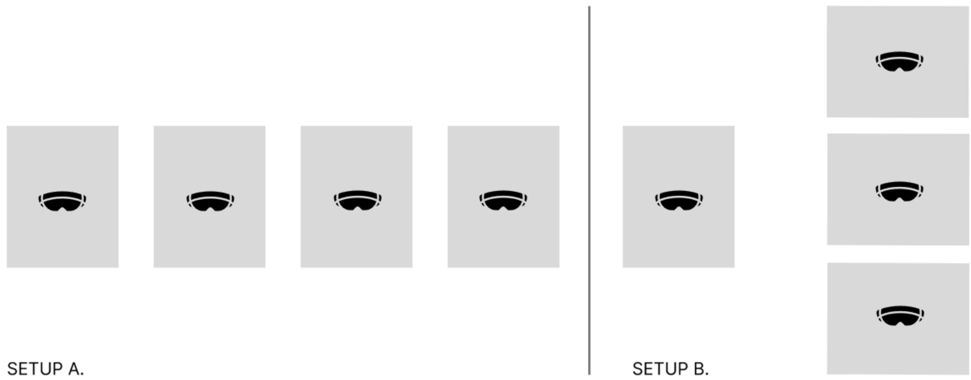


Figure 25. The setups for testing. Setup A was used in the first lesson and setup B in the second lesson.

The students' feedback questionnaire (see Appendix 2) is a shorter version of teacher's questionnaire and the themes studied are similar (i.e., background questions, experience, usefulness, usability and ease of use and VR in school) as shown in Figure 26. However, to the final theme was added questions regarding suggestions for improvement and features that were or were not liked. The questionnaire was anonymous similarly as the teacher's questionnaire was. The questions utilized were open- and closed-ended questions, and Likert scale question sets were utilized as well. The scales used in both questionnaires were similar. To facilitate the collection of responses, a printed paper version of the questionnaire was used with students.

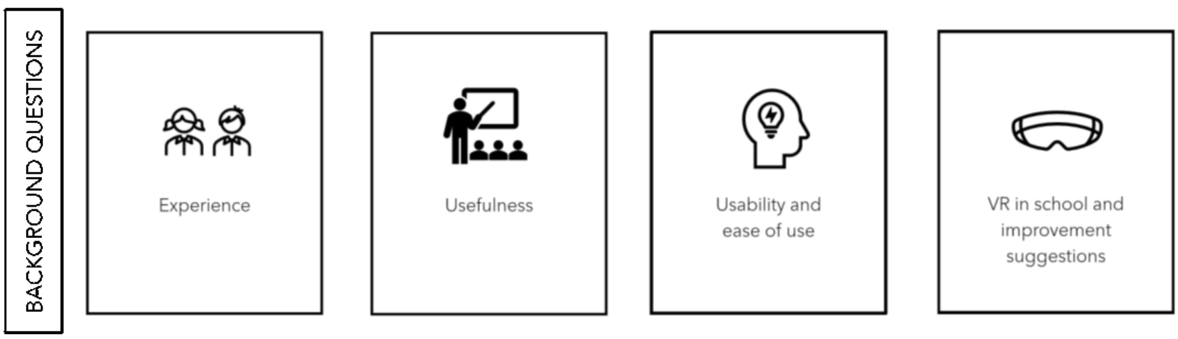


Figure 26. Themes of the student's feedback questionnaire.

The questionnaire started with a short introductory to inform what the questionnaire is about, and for example, how answers were handled. The background questions solved the students' interest regarding learning biology (3A) and the interest of using the VR headset (3B. –C.) as well as the prior experience of VR. Experience (4) related questions discovered matters related to user's experience such as liking the game, having fun, the safety aspect and the experiences of feeling nausea while tested. Usefulness (5) theme discovered the learning aspect and the awareness of the progress during testing and the feedback features the game offered. Usability and ease of use (7) related to questions solving how easy it was to start using the VR headset, were there enough instructions, what an experience the use of hand controllers was and did students feel they could influence the environment. Multiple statements existed in the students' feedback questionnaire for each requirement R1–R12 that teachers set as shown in Table 12. Open-ended questions of the theme aimed to offer more details and discover, for example, what students found difficult to do and on the other hand, what was easy and effortless to perform. The final set of questions related to VR in school and improvement suggestions. Questions discovered, for example, what is the frequency students prefer to have VR sessions, how can the game be made more comprehensive, what they liked and did not, and what they would prefer to learn in school with VR. The final open-ended question offered an opportunity for students to bring out and highlight something related to testing.

6.2 Analysis and results of user testing

The collected data was studied to discover students' preferences and experience related to the VR learning material and to answer research questions RQ 1: "What are the main components of valuable VR learning material?", RQ 4: "Is there a difference between teacher's and student's VR preferences?", RQ 5: "How can VR material be implemented into education?", and RQ 6: "What are the challenges of VR learning?". First, the background questions were explored, to identify the students' prior experience of the use of VR and the interest related to the subject of biology. After that, the results of the teachers' and the students' questionnaire were compared in two different ways. First, the most agreed statements and themes emerged in students' testing were solved. Then, it was clarified how the requirements set by teachers were met in the application developed. After that, each quality requirement was investigated considering all the data collected to clarify problems, details and to figure out possible solutions. The observational data was collected from 17 students who participated in the user tests. 14 of the students returned the feedback questionnaire and these answers are employed in descriptive statistics and in calculations. The frequency that teachers and students prefer VR lessons should be organized is compared at the end of this chapter. Some modifications are made before the calculations and analysis are enabled and the students' data is comparable to the data of teachers.

6.2.1 Modifications and background questions

For the Likert scale used in the students' questionnaire, some adjustments needed to be made. The questionnaire yielded seven "Undecided" answers and moreover three students had understood the scale as continuous-scale and given values between two scale points. Ten replies were placed in the middle of the scale (i.e., between somewhat disagree and somewhat agree values). To make all replies comparable, the scale was modified similarly as in the teacher's questionnaire (i.e., 1 to 5). Undecided values and the middle scale replies correspond the value three after modification. Another modification made was to rescale statements 4.G. and 7.D. the other way around. The statements received are 4.G. "I did not feel nauseous" and 7.D. "I did not get stuck and know how to proceed".

The reliability of statement groups related to quality requirements (see Table 12) were studied to enable the comparison of requirements later, and after, ten statements (gray and not bolded) were removed in order to obtain a consistent and representative measurement for each requirement. The size of groups was one to seven statements and similarly as in teachers' data analyzed the larger the group size, the more data is involved. A large amount of data can be more comprehensively utilized than a smaller amount. Most of the groups studied were relatively small and for this reason, smaller reliability values are accepted for normal.

Table 12. Quality requirements R1–R12 with corresponding statements of the students' questionnaire and reliability value. Removed statements after reliability analysis are grey and not bolded. Note k = number of statements.

R1	4A, **4B**, 4D, **4E, 5A–C, 7B, 7C,** 7D, 7F (Cronbach's alpha 0.73, k = 7)
R2	**7B, 7C, 5C,** 7D (Cronbach's alpha 0.60, k = 3)
R3	**5A, 5B** (Cronbach's alpha 0.84, k = 2)
R4	**4E** (k=1)
R5	**4E, 7F** (Cronbach's alpha 0.80, k = 2)
R6	**4B, 4E** (Cronbach's alpha 0.55, k = 2)
R7	**4A,** 4B, **4F** (Cronbach's alpha 0.78, k = 2)
R8	**4E** (k = 1)
R9	**4A, 4C, 4F** (Cronbach's alpha 0.78, k = 2)
R10	5C, **7A**, 7C, 7D, 7E (k = 1)
R11	**4D** (k = 1)
R12	**4G** (k = 1)

Background questions

Background questions solved students' attitudes to biology learning and the experience of using VR headsets. The results can be considered representing more on students' perceptions of themes than the actual case since the questions asked were very straightforward. This approach was chosen due to interest in the overall opinions and attitudes of students and on the other hand to keep the questionnaire compact and easy to answer. Because of the small sample size (n = 14) and the compactness of the questionnaire statistical significance is not calculated in this case.

Examination revealed that half of the students (n = 7) had prior experience in using VR headsets. Most of the students did not remember the model of the headset they had used, but one reported that the model had been a cardboard model. Some students reported having used VR headsets at school. For that reason, it can be assumed that all students had used the smartphone requiring models such as a cardboard model and had very little experience of the kind of VR headsets they were about to test in the study that can be interacted with hand controllers. According to background questions students' interest in learning biology was a bit lower (see Table 13; M=3.57 (3A)) than the interest of using the VR headset (before testing M = 4.29 (3B), after testing M = 4.36 (3C)) and the mean difference which measures the difference of two means [228] was MD = 0.72. Students reported that their interest in the device slightly increased (MD = 0.07) during testing.

Table 13. Descriptive statistics of background (3) and experience (4) related questions.

	3A	3B	3C	4A	4B	4C	4D	4E	4F	4G
Min	1	2	3	3	2	1	2	2	3	1
Max	5	5	5	5	5	5	5	5	5	5
Mean	3.57	4.29	4.36	4.5	4.57	4.07	4.36	4.43	4.5	3.5
Median	4	5	4	5	5	4	5	5	5	4
Std. Deviation	1.45	1.07	0.63	0.65	0.85	1.07	1.01	0.85	0.65	1.70

Table 14. Descriptive statistics of usefulness (5) and usability and ease of use (7) related questions.

	5A	5B	5C	7A	7B	7C	7D	7E	7F
Min	2	2	1	3	1	2	1	2	2
Max	5	5	5	5	5	5	5	5	5
Mean	3.57	3.79	2.93	4.57	3.86	4.21	2.57	3.93	4.36
Median	4	4	2.5	5	4	4.5	2	4	5
Std. Deviation	1.28	1.12	1.27	0.76	1.23	0.98	1.56	0.92	0.93

6.2.2 The success of the requirements' implementation

It is compared whether the same themes exist in high rated statements both in teachers' and in students' questionnaires. The focus is to study whether the statements and teacher's requirements are successfully implemented into the application. First, the mean values for each theme are calculated in Likert scale question sets (i.e., the experience, the usefulness and the usability and ease of use) as shown in Table 13 and Table 14. Statements were similarly divided into four categories as in the teacher's questionnaire. The success of implementation is evaluated by comparing the results of the teachers' questionnaire and the students' testing. The students' feedback questionnaire is studied, and the statements are categorized to recognize students' opinions. The scale used is: "very strongly agree" ($M \geq 4.5$), "agree" ($3.5 \leq M < 4.5$), "not agree nor disagree" ($2.5 \leq M < 3.5$), "disagree" ($1.5 \leq M < 2.5$) and "very strongly disagree" ($M < 1.5$).

It is discovered that four statements are "very strongly agreed" and have value $M \geq 4.5$ as shown in Table 15. When the reliability of the group is studied, it is found that 4.B. has a negative item-total correlation value (-0.03). However, because of a small sample size the statement 4.B. is decided to include the final set (Cronbach's alpha 0.56, n = 4). The four most agreed statements are listed below, and after each statement, the related requirements are introduced. Three of the four statements that students agreed with the most (75%) relate to students' experience (4A, 4B, and 4F). The final statement (7A) relates to ease of use (25%). After the questions related to educational content, which can be problematic since they cover many themes, are removed the ration is very similar to teachers' questionnaire (i.e., students' experience 67 % and the ease of use 11 % of statements were considered very important).

Table 15. The group of "very strongly agree" related statements in the students' questionnaire.

4.A. I liked the VR game (Relates to R1, R7, R9)
4.B. Using VR-glasses provided pleasant change to learning (R1, R6, R7)
4.F. I had fun (R1, R7, R9)
7.A. It was easy to start using VR-glasses (R10)

The "agree" group ($3.5 \leq M < 4.5$) include nine statements as presented in Table 16. While studying the reliability of the group it was discovered that 4.D. had the item-total correlation value 0.06 which is lower than the value that is normally excepted (< 0.2). In this case, however, the statement is included in the final set to obtain larger amount of data. 22 % of statements in the agree group related to the experience (4C, 4E), 11 % to the safety (4D), 22 % to the usefulness (5A, 5B) and 45 % to usability and ease of use (7A, 7B, 7E, 7F).

Table 16. The group “agree” related statements in the students’ questionnaire.

4.C. I liked the environment of the game (R9)
4.D. I felt safe to use the application (R1, R11)
4.E. I felt encouraged to try new things (R1, R4, R5, R6, R8)
5.A. The VR-game helped me to learn new species (R1, R3)
5.B. I learned new things (R1, R3)
7.B. I received instructions enough from the game (R1, R2, R10)
7.C. It helped to hear instructions from others (R1, R2, R10)
7.E. Using touch controllers was easy (R10)
7.F. I was able to performs actions I wanted in the game (R1, R5)

Three final statements belong to the category “not agree nor disagree” ($2.5 \leq M < 3.5$) as shown in Table 17. The final group has a negative Cronbach’s alpha value -0.328 since it has a negative average covariance between statements. Negative Cronbach’s alpha value usually is due that some items behave in reverse. It was not possible to receive acceptable reliability value with these three statements. The group is studied as a whole in this case based on mean values, even though the group is statistically heterogenous. However, it is recognized that the mean value does not necessarily indicate how the claims have been answered. One statement of the group (4G.), relates to safety (33%) and two statements (5C, 7D) to the usability and ease of use (66%).

Table 17. The group “not agree nor disagree” related statements in the students’ questionnaire.

4.G. I did not feel nauseous (R12)
5.C. I received enough feedback from the application (R1, R2, R10)
7.D. I did not get stuck and know how to proceed (R1, R2, R10)

There were no statements to put into the two lowest categories (i.e., somewhat disagree or strongly disagree). Thus, it can be considered, that the most of the quality requirements (n = 10) implemented into the application were somewhat successfully implemented and discovered by students.

The comparison of teachers’ expectations and students’ testing feedback

Next, each requirement with its group of statements were compared. The value of students’ feedback is received by calculating mean values of the statements related to each requirement R1–R12 (see 6.2.1.). Mean values are compared to values obtained from the teachers’ questionnaire to solve whether the implementation of requirements was successful. To categorize the success of implementation, instead of using Nielsen's severity rating which is developed to evaluate the UIs in the first place, a scale is used based on the idea of Levy and Novak's (2010) bug severity levels which is useful for evaluating games. A bug is something that is not planned to exist in an application but occurs nonetheless (i.e., flaw) and is sought to locate and repair as early as possible in the development process. Bugs can relate to, for example, visuals, audio, level design, artificial intelligence (AI), physics, stability, performance, network, or compatibility of the game. The bug severity level is a scale to evaluate qualities and the severity of each bug found. The scale is a four-point scale and has the categories: low, medium, high, and critical. [255] For example, a bug that makes the application crash is a critical bug. In this study, a fixed numerical scale is used, which is received from mean differences of teachers' expectations and the students' feedback, and bugs are considered in a wider aspect and not as single separate problems. It is decided to make the scale quite tight and with short scale intervals since it is known that a novelty factor exists at the beginning when starting to use new digital items [118], and students probably tend to give more positive evaluations at that point than in a normal situation. In the classification utilized in this study, the category low is replaced with the succeeded category.

From the graph Figure 27 and Table 18 below can be seen that teachers' expectation values are all a bit higher than the feedback values given by students. Requirements that have MD (mean difference) ≤ 0.5 the implementation is considered to be succeeded. 83% ($n = 6$) of the quality requirements that are successfully implemented, relate to the experience theme (R5 to R9) and 17% to the usability and ease of use (R10). The result is very similar to the comparison made in the first examination (i.e., separate statements) The second category with medium priority ($1 \leq MD < 0.5$) includes features which need consideration and "should be fixed soon" ($n = 4$). Requirements relate to the usefulness (R1, 25%), the usability and ease of use (R2, 25%), the experience (R4, 25%), and the safety and health (R11, 25%). The third category with high priority is arguing the implementation "must be fixed soon" ($1.5 \leq MD < 1$). One usefulness requirement R3 "Material intensifies the knowledge of students" relates to this category ($n = 1$). In fourth category and with the mean difference higher than 1.5 the repair is critical to perform and "need immediate attention" (critical priority). One safety related requirement R12 "Moving does not cause motion sickness", includes this category ($n = 1$).

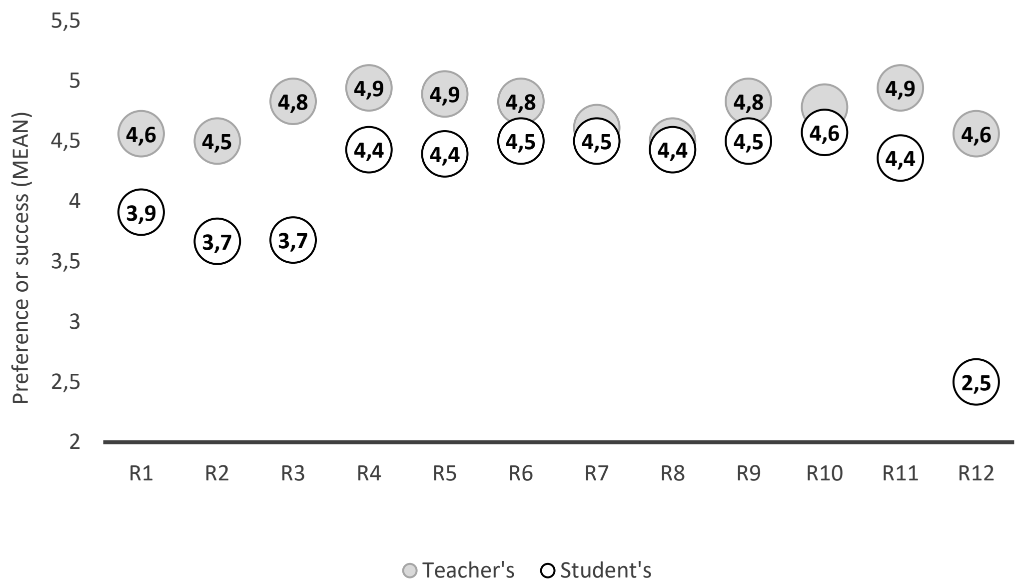


Figure 27. Comparison of mean values of Teachers' expectations and success of implementation.

Table 18. Descriptive statistics of statements related to quality requirements of student testing.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Min	3	2	2	2	2	3	3	2	3	3	2	1
Max	5	5	5	5	5	5	5	5	5	5	5	5
Mean	3.91	3.67	3.68	4.43	4.39	4.50	4.50	4.43	4.50	4.57	4.36	2.50
Median	4	4	4	5	4,5	4,75	4,5	4	4,5	5	5	2
Std. Deviation	0.68	0.87	1.12	0.85	0.81	0.71	0.59	0.85	0.59	0.76	1.01	1.70
Teachers' Mean	4.56	4.5	4.83	4.94	4.89	4.83	4.61	4.5	4.83	4.78	4.94	4.56
MD	0.65	0.83	1.15	0.51	**0.50**	**0.33**	**0.11**	**0.07**	**0.33**	**0.21**	0.58	2.06

Detailed report of the success of implementation of requirements

In the quality requirement Table 19 below, is combined all data available and collected in the study. It details the success of implementation of requirements and the feedback received from the testing. The succeeded requirements are studied all together, but other requirements are separately examined. In the Table 19 the priority of bugs, themes, related details, and the proposed solution are introduced. The statements, earlier removed because of inadequacies of reliability are reintroduced in this examination and utilized in the reasoning. In some examination of requirements, graphs are attached to demonstrate findings. Graphs are based on open-ended questions of the student's feedback questionnaire, and the number of responses vary especially in students' case since answering questions was optional.

Table 19. The quality requirements table includes detailed information about the features succeeded, and the bugs found which related to problems in implementing teachers' preferences.

R1.	Material obeys the curriculum (Uf; 4A, **4B**, 4D, **4E, 5A, 5B, 5C, 7B, 7C**, 7D, 7F)
Priority:	Medium (MD = 0.65)
Theme:	Usefulness
Problem:	Not following the curriculum enough.
Details:	The first requirement R1 can be considered as an umbrella requirement covering statements used in the most other requirements as well. Students somewhat agreed that the material obeys the curriculum (M = 3.9). Some of the statements included the requirement were more agreed than others.
Possible reason:	Testing time was too short for learning purposes especially when students were using the device for first time. And as said earlier, measuring learning is not reasonable in this test setting.
Suggested solution:	Give more time and more instructions to make more efficient use of the time available.

R2.	Help is provided when needed (UfUb; **5C, 7B, 7C,** 7D)
Priority:	Medium (MD = 0.83)
Theme:	usefulness, usability
Problem:	Students got stuck with the environment and did not know what to do.
Details:	The first-time experiences of the use of VR headset are highlighted in testing and results. Since the device and the environment are both new, more instructions are expected. Most of the students considered having enough instructions from the game. Two students did not agree and another strongly disagreed. With statement 7D four students strongly agreed "getting stuck", and five students somewhat agreed on this. But after all, there were no students saying that it would have been difficult to start using the VR headset or saying that they did not have fun or not liked the game. Students reported that difficult had been taking and holding items in hand and catching species 64 % as seen in Figure 28 above. Understanding instructions at the beginning was another thing mentioned (27%). When students were asked to name features to make the game more comprehensive, 58% named instructions as illustrated in Figure 28 below.
Suggested solution:	To add extra guidance in case student's do not figure out what to do and generally more and more precise instructions (e.g., visuals (e.g., DOF) and spatial audio [200]). One student accidentally teleported before reading instructions and figuring out what to do became a struggle. How to guide users to look at the right place in VR has been seen as a challenge also earlier [178]. To disable movement and to enable, after instructions are read is something that could help. Another student suggested that there could be 3D-models of hand controllers in the environment to help to understand how hand controllers work. A bigger environment was requested, and it could prevent students e.g., throwing and losing their equipment cross the worlds boundary. It was also suggested that lost items could teleport back to their original place.

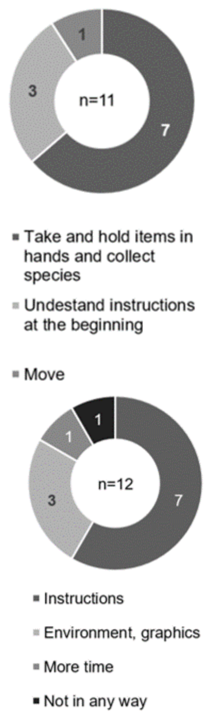


Figure 28. What was difficult (above) and how to improve (below).

R3.	Material intensifies the knowledge of students (Uf; **5A, 5B**)
Priority:	High (MD = 1.15)
Theme:	Usefulness
Problem:	Students did not learn species.
Details:	Five students admitted that they somewhat disagreed learning new species during testing, and the rest of the students strongly agreed or somewhat agreed learning species. However, only two students were able to name species correctly. One student remembered two species, a dragonfly and a dragonfly larva, and another swan mussel. Other answers received were a larva, an insect, a worm, or nothing. Other features than species that students remembered were e.g., catching equipment (n = 8) or things related to the environment e.g., trees and water (n = 5) (see Figure 29).
Possible reason:	Not enough time and/or instructions available.
Suggested solution:	Give more time and add instructions. The learning experience could be enhanced by adding 3D-models of species.

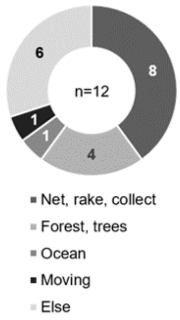


Figure 29. What else students remembered.

R4.	Students experience things instead of standing by and watching (EUf; **4E**)
Priority:	Medium (MD = 0.51)
Theme:	Experience, Usefulness
Problem:	Students did not know what to do and wandered around the environment.
Details:	The mean of the requirement was quite high (M = 4.4). One student somewhat disagreed, eight students strongly agreed and the rest of the students somewhat agreed. However, the value of teachers was expecting was even higher. The mean difference value is even though very close to be interpreted succeeded implementation (MD = - 0.01).
Suggested solution:	Adding more and more precise instructions to enhance the agency of students. Giving more time can help as well. One student mentioned it disturbed the experience, to see the real environment from the gap above nose and was hoping to have something to cover that.

R5.	Student can explore environment (EUf **4E, 7F**)
R6.	Student gains new experiences (EUf; **4B, 4E**)
R7.	The motivation to learn the subject increases. (EUf; **4A,** 4B, **4F**)
R8.	Student can experience something that is not possible in real life e.g., fly (E; **4E**)
R9.	Student will get positive learning experience (EUf; **4A, 4F**)
R10	Application is easy to use (Ub; 5C, **7A,** 7C, 7D, 7E)
Priority:	Succeeded (MD = 0.07–0.50)
Theme:	experience, usefulness, usability
Details:	One or two students were undecided or somewhat disagreed, and the rest of students agreed or strongly agreed of these six requirements categorized succeeded. 54% of students considered that they most succeeded with teleporting and moving (n=7, see Figure 30 on the left side). The best in VR was named to be freedom and moving around as illustrated in Figure 30 on the right side.

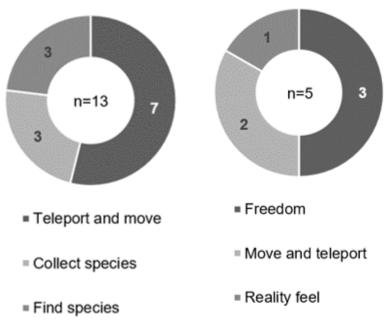


Figure 30. What succeeded (left) and the best in VR (right)

R11.	It is safe to use (SUf; **4D**)
Priority:	Medium (MD = 0.58)
Theme:	Safety, Usefulness
Problem:	Some students felt the environment unsafe.
Details:	There was one student who told somewhat disagreeing with the feeling of safety during testing. Two other students were undecided about the requirement. One possible reason for not feeling safe was an issue found an open-ended question which solved what was unpleasant. Students named feeling nausea (n = 4) and experiencing headache or migraine (n = 3, see Figure 31).
Suggested solution:	Make sure the guardian of the headset is on, and there are no outer distractions. Add instructions to raise the comfort. Fix problems related to nausea and migraine. For example, remove problems related to shaders and movement related issues. One student mentioned that VR glasses could be lighter which also would possibly make the use more comfortable.

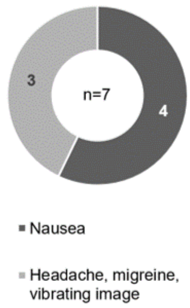


Figure 31. What was unpleasant.

R12.	Moving does not cause motion sickness. (S; **4G**)
Priority:	Critical (MD = 2.06)
Theme:	Safety
Problem:	Students experienced nausea when exploring the environment.
Details:	The mean difference was the biggest of all (MD = 2.06). Teachers named the requirement to be very important (M = 4.6), but many students experienced nausea (M = 2.5). Average student "did not agree nor disagree" feeling nausea. Six students strongly disagreed, and three students strongly agreed not experiencing motion sickness during testing. The issue polarized students strongly. Similar result was received from an open-ended question where four students reported that feeling nausea had been the most unpleasant feature with VR. (see Figure 31)
Suggested solution:	To make sure students have adjusted the stripes of the headset in the best possible way and the pupil distance is set right as well. Reduce height differences from the virtual environment for smooth transition. Eliminate the unnecessary movement with more detailed instructions. Enable teleporting longer distances. Consider adding vignette effect. Give more time so that students do not need to rush and/or disable fast movements.

6.2.3 How to implement into education

To answer the final research question RQ 5: "How can VR material be implemented into education?" students were asked what themes they would prefer to learn with VR and how often VR lessons should be organized in schools. It was discovered that with VR the students (n=7) preferred to learn biology and science (29%) and physical education e.g., football (29%). The majority of students were not able to name a specific subject or theme but reported preferring to learn everything (42%) with VR. Students are different and some are bolder, for example. One student mentioned willing to try out a roller coaster with VR and on the other hand another student hesitated to go into the virtual water. According to VandenBerghe (2012) players can be divided into five groups based on their qualities and what kind of games they prefer. The features that other players seek, and others avoid are novelty, challenge, stimulation, harmony, and threat. [198] [263] [264]. To design an application that meets the needs and is considered interesting for the variety of students is a challenge.

While playing games, different roles are obtained by players. Considering the goals of players, it is discovered eight different roles. Griefers want to dominate others and to have bad reputation, opportunists seize every opportunity, hackers try to find new phenomena, friends enjoys familiar company, politicians seeks for a big and good impact, planners try to achieve their own goals, scientists explore the game to identify how it works and networkers interact with people they do not know beforehand [265]. At least a couple of hacker-type players were found in the class tested since some students discovered that throwing equipment in the emptiness is a way to lose them. Other player types existed too, and analyzing this would have been interesting, but within the limits of this study, this was not further explored.

The second question that students were asked was the frequency of VR lessons they preferred. Students were some amounts more eager than teachers to use VR in school and 50% of students considered that VR should be included every or every third lesson. However, not every student was agreeing on this. One student said that VR lessons should not be arranged at schools at all, and two other students were preferring the use of VR one or a couple of times during the school year. In the testing phase, one student also needed to terminate the testing early. The possible reason for this may have been based on the overall feedback: the nausea, migraine or headache, but weather there was a specific reason for interrupting the testing was not revealed in the testing.

Finding out RQ 4: “Is there a difference between teacher’s and student’s VR preferences?” a final comparison was made. The difference was evaluated with Mann-Whitney U-test. The test was used after the data of student’s is not normally distributed and samples are independent. In the analysis was found a significant difference ($p=0.028$) between students’ and teachers’ preferences. The results are presented in Table 20, and it revealed that students ($M = 2.71$, $SD = 1.49$) felt that VR sessions could be arranged significantly more often than teachers considered ($M = 3.50$, $SD = 0. 71$).

Table 20. Descriptive statistics about teacher's (n = 18) and student's (n = 14) perception of the frequency of VR session.

	Students	Teachers
Mean	2.7	3.5
Median	2.5	3.5
Std. Deviation	1.49	0.71
Mann-Whitney U	70.5	
Wilcoxon W	175.5	
Z	-2.2	
Asymp. Sig. (2-tailed)	**.028***	
r	-0.39	

* = significant difference.

7 Discussion

7.1 Limitations

The limitations of this study relate to testing equipment, the testing situation, testers, modifications made for the scales and sampling. The first limitation examined here is the limited number of test equipment. After the research permit, approved by supervisor, had been accepted by school to be tested, it appeared that there were difficulties to acquire headsets for testing. For later studies, it is recommended to double check the availability of test equipment and to discover the existing practises and the expectations that obtaining the equipment will require. Since even at the end, it was not possible to acquire enough VR headsets, the testing was performed with four headsets instead of ten. Heavily modifying the test setting was not reasonable at that point, so the testing time per student was significantly reduced.

The alteration impacted not only the testing situation, but the feedback questionnaire as well. The questionnaire had to be made very compact, quick and easy to answer. Even though some students may have still rushed filling the feedback which can have an impact on the results as well. One student mentioned in the feedback that giving more time for the testing would have enhanced the VR experience. In previous studies it has been discovered, that the design process can be biased, for example, because of limitations of resources or lack of time [97] and unfortunately both of these elements were present in this study.

An issue that All et al. (2016) have observed, to impact on results is having an external instructor guiding a test situation. However, since the school tested is one of the university's training schools, students are very familiar with working with different teachers and student teachers. In addition, the person that guided the session is a biology teacher as well (i.e., a double specialist) which increased the comfort level and the experience of an authentic classroom situation. Thus, having an unfamiliar person guiding the session can be considered having a minimal impact on the environmental conditions and final results.

The third factor which can have an effect on the results is that the tested class had S2 (Finnish as a second language) students too. S2 students have weaknesses in basic language skills [266] and understanding the questions of the questionnaire and instructions of the application may have been more challenging. However, all students tested followed a normal biology syllabus and studied in a science-oriented line.

In addition, of possible difficulties of understanding the language, a confusion was also with at least one statement related to the usability (5C.) ”I received enough feedback from the application” (R1, R2, R10). It appeared that some students did not know what was meant with the statement, but since clarification was asked, it caused no issue. However, not all students are willing to solve problems that occurs, and alternative interpretations may still have existed. An example would have helped students to understand, and if more time had been available, there could have been multiple and more precise statements related to that theme.

Considering the analysis, some modifications needed to be performed. The modifications made for the Likert scale used in questionnaires may have had some impacts on the final results. Placing the last “undecided” value in the middle of the scale to represent the value 3 may be considered in some cases problematic, and can cause distortion. Some teachers may have interpreted this value as “I do not want to say”. However, since the original middle value was missing, most of the teachers probably considered the value as a neutral value and the replacement caused no impact on the results. If the modification had not been made, the missing values would probably have caused more harm. There were also some themes that had overlapping statements with other themes that may have impact on results, lowered significance levels and made measures more similar. Overall, the MANOVA analysis had twelve overlap statements, and the degree of overlap was from 16 % to 67% (M = 44%) of all statements of themes as seen in Table 3. However, it is difficult to find measures that would represent only one theme since in learning many things are connected with each other. Curriculum covers not only the subject, but many other themes as well. For example, seven of the statements measuring usefulness relates to usability and ease of use, two to safety and health, three to teacher’s role and three to student’s experience as well (see Table 2).

The whole data collected consists of teachers’ questionnaire (n = 18), students’ user testing (n = 17) and students’ feedback questionnaire (n = 14). Because of the relative small sample size, smaller reliability values are accepted, for example, in the evaluation of how teaching experience impacts on preferred features. For example, the sample size impacts the reliability of the r value through distortion of the z value [267]. Another thing is that with more statements involved in questionnaires there would have been less need for removing statements in reliability examinations. With bigger data amount reliability values are higher.

Cohen’s Power tables reveal that the significance level of 0.05, n = 18 and medium effect size (r = 0.3) is expected to receive the power of 13. To receive the power of 80 that in many cases is aimed at requires, for example, effect size of 0.70 or 110 participants (r = 0.3) [227]. All et al. (2016) have considered that the sample size should not be less than 20 and, if statistical analysis is run, 30 is the minimum value of participants per condition [56]. In this study, the study setting and smaller sample size were accepted for practical reasons and because of limited time resources. Results received from this study may not be generalizable because of the use of non-probability sampling and the small sample size. Even though the results of this study can be considered indicative and provide interesting point of view and new insights for the theme of VR learning and learning material design.

7.2 Analysis of findings

In this analysis of findings, the research questions RQ1 – RQ6, and the implemented requirements that were categorized succeeded or need fixing are discussed. Some environmental factors that rose from the study are examined also. After the challenges and issues that can hinder learning (RQ5) are presented, possible solutions are discussed in the light of earlier research findings.

Table 21. The success of implementation of quality requirements (R1 – R12).

Succeeded	R5.	Student can explore the environment (EUf; MD = 0.50).
	R6.	Student gains new experiences (EUf; MD = 0.33).
	R7.	The motivation to learn the subject increases (EUf; MD = 0.11).
	R8.	Student can experience something that is not possible in real life (E; MD = 0.07).
	R9.	Student will get positive learning experience (EUf; MD = 0.33).
	R10.	Application is easy to use (Ub; MD = 0.21).
Medium	R1.	Material obeys the curriculum (Uf; MD = 0.65).
	R2.	Help is provided when needed (UfUb; MD = 0.83).
	R4.	Students experience things instead of standing by and watching (EUf; MD = 0.51).
	R11.	It is safe to use (SUf; MD = 0.58)).
High	R3.	Material intensifies the knowledge of students (Uf; MD = 1.15).
Critical	R12.	Moving does not cause motion sickness (S; MD = 2.06).

Most of the requirements teachers set and found very important related to the students' experience or the usefulness theme. The implementation of requirements in the learning material was succeeded for half of the quality requirements (R5 – R10). Succeeded requirements related mainly to students' experience (n = 5) and usefulness (n = 4), but one requirement related to usability (n = 1) theme as well (see Table 21). Four requirements (R1, R2, R4, R11) need to be fixed with medium priority, one with high priority (R3), and one was found critical to repair (R12). Requirements that needed fixing connected mainly to the usefulness (n = 5), but also the safety (n = 2), the student's experience (n = 1) and the usability (n = 1) theme. With the number of statements needed to be repaired must be remembered that the test setting was no favoring to measure subject related learning which impacts the students' experience of usefulness of the application.

The very strongly agreed statements among students connected to their positive experience of VR and the ease of getting started to use devices. The feeling of freedom was something students liked in VR learning the most, and they considered that the easiest part of the VR environment was to move. In general, what students are allowed to do and where to go is very restricted since responsibilities in the school. VR technology can offer students a simulated, realistic place or another world inside school with more freedom but simultaneously be safe, under control and supervised. Even though students reported the easiness of starting the use of VR the least agreed statements related to instructions and feeling nausea. It revealed that the features students raised up in their feedback as features to be improved are the same issues teachers had named very important (i.e., requirements), but the implementation had not succeeded. Students argued, for example, preferring wanting for more instructions, feeling nausea and not having enough time. They did not say directly, that they could not learn in the VR environment, but named elements that made learning more difficult, and on the other hand features that would help them to learn. This implies the importance they saw considering learning and the usefulness of the material.

7.2.1 Learning conditions and environmental resources

Alalwan et al. (2020) have studied difficulties that are faced in VR learning and found four main challenges that are experienced. Environmental resources are an issue found problematic from teachers' perspective, for example, the lack of devices, a poor internet connection, or not having enough space in the classroom. Meta Quest 2 headsets were found useful in the testing of this study, and suitable for the purpose. Devices were recharged during breaks and had enough battery capacity to carry through testing. The standalone feature of headsets was practical and enabled testing and usage in different places, also in corridors. In this study, 78% of teachers reported that they would prefer to have devices at least for every other student if not every student for effective learning. The effectiveness of learning has been discovered being important in previous studies as well [268]. Unfortunately, how to execute this in the classroom space could not be tested because of the lack of devices. More headsets would be needed to enable an authentic test situation.

Providing students a fairly large play area was a challenge, even with four headsets and setting up the classroom in advance for the VR testing needed some time. It is expected that more challenges and more time for preparation would been needed if more headsets were available. It was noted that classrooms have very little extra space for movement and using, for example, corridors can be advantageous but, spaces outside the classroom can have distractions.

Encouraging and to get students to move more is a new goal of the Basic Education Act which goes into effect next year in the schools of Finland [269]. Most of the students tested moved quite a lot during testing, and some also jumped while trying to catch species flying high. Getting students to stand and to move is one advantage of VR learning since there can be a lot of sitting in a classroom generally. However, in many cases with VR a stationary boundary can come into question because of the lack of space, in case corridor space or gym are not available to use. At least with longer sessions, the possibility of sitting can be considered useful for practical reasons, and it can be beneficial to design places where students can sit and rest in the virtual environment as well.

In the Alalwan's et al. (2020) study, the second most common challenge for teachers with VR was the lack of time and practice [270]. It was found that in this study teachers with over ten years teaching experience preferred to have significantly more help and instructions from the VR application and a less active role for themselves. With more comprehensive instructions, it can be considered, that teachers as well need less time to preparation and practicing, which solves the lack of time issue at some level. It is also students' desire not needing to ask for help from the teacher or others since they reported preferring more comprehensive instructions from the application as well.

In schools and in the context of education the curriculum and the learning goals do set a very strict schedule for things that need to be learned during the school year even though the teacher has a quite wide autonomy, for example, to choose tools and materials to utilize [93]. Thus, for a single theme cannot be used too much time and if more time is required more learning goals are expected to be met. The effectiveness requirement conflicts with the original idea of games where performance is not something to seek for [28] however in serious games gamifying elements can be added without losing the effectiveness.

Teachers of this study considered that the VR session can be from 11 minutes to 40 minutes, and designing material that is flexible in terms of time is beneficial. The time teachers preferred is much less than the two-hour limit that Villena-Taranilla et al. (2022) have set as maximum length of VR sessions in K-6 education [24]. But of course, the time-consumption depends on the material, the task students need to deliver, and how familiar they are with the devices already. To allocate at least 20 minutes per student for usage in case they already are familiar with devices and at least half an hour if they are not, might be beneficial. However, after starting the use, students expect to be able to play for a while. For example, in student testing because of the time limitation, testing needed to be interrupted very quickly after the start. For most students, a couple of reminders were needed after they were ready to return the devices. Adding an element where the teacher can end the game is something to consider. The feature could be, for example, similar as in the video meetings' breakout rooms with a count-down before all participants return into the same space [271]. Also, some kind of animation or more gradual dislocation while returning back to the real classroom could be beneficial and reduce the unpleasantness of just removing the headset.

7.2.2 Instructions and help

In 58% of the feedback received from the students, they asked to have more and more comprehensive instructions. This relates to R2. "Help is provided when needed" and R4 "Student experience things instead of standing by and watching". Both requirements are categorized to medium priority defects. More comprehensive instructions make learning more effective. There can be multiple ways to implement instructions as discussed earlier [236] [237] but in the simulation style environment adding a scientist or a fellow student to guide could be a useful option and would not interrupt the experience that much. Somehow highlighting equipment and the species to catch could be one possible solution and help students to figure out what to do. Designing a learning material that balances between having enough instructions and the inquiry learning approach that was chosen for the material is a challenge.

Adding a tutorial at the beginning of the material could help students to start the use of the device, and to familiarize the environment. This was something teachers saw important ($M = 4.4$) and was very close to becoming a requirement (≥ 4.5). Restricting movement before instructions are read is one option, however it can reduce user control which is seen important [242] but might be in this case justified. Adding real resampling 3D-models of hand controllers in the VR environment with labels can ease the use especially at the beginning. 3D-models of species can improve the experience. To add modalities one possible option is adding signifiers and sounds, for example, for insects that are aimed to be caught to tempt users to interact with them. For some students, it might be helpful also to hear advice, and adding sounds can improve the 3D-experience. [12] [16] Scaffolding during the game play session is also important and, for example, adding a code for lost items that return those in their original place (proposed by a student) can prevent students from getting into a situation where they can no longer progress in the game. It is important that students have all items needed and losing key items during the game play is prevented.

Overall, it was discovered that the students' interest using VR devices increased slightly during the testing. It would have been interesting to ask students whether the interest to learn biology remained the same or increased similarly. This is something that can be considered examining in further studies. Making more precise questions on what students actually learned and covering other themes not only the subject can be beneficial in future studies as well. The most students considered that "It was easy to start using VR glasses" (R10 / 7A). However, especially one student (A) struggled to use the headset and ended up choosing the undecided option for the three most agreed statements. The student also reported that he/she was strongly agreeing on feeling nausea, which may explain the student's responses which differ from the average opinion of the class. This implies that nausea can have a great impact on overall experience and other tested factors as well for an individual student.

7.2.3 Nausea and safety

Students reported in an open-ended question that the worst part with VR was feeling nausea (57%) and experiencing headache, migraine or that the image was vibrating (42%). It was observed that the implementation of R12. "Moving does not cause motion sickness" had not succeeded as well as other requirements and students reported "not agreeing nor disagreeing" feeling nausea (M = 2.5). This was categorized as a critical priority defect (MD = 2.06) that must be fixed soon. Even though there may have been some distortion related to the statement 4G. "I felt nausea" since other statements of the group were positive and the last one had a negative stigma the nausea experience cannot be considered being just a statistical illusion after students had also written about the phenomenon in an open-ended question as well. In a previous study, it has been discovered that with non-gamers the experience of cybersickness is usually higher [272] which may have an impact on the results of this study as well. However, if the first experience with VR is not good, that may impact on how willing students are to use the technology for a second time. For example, in this study, one student who experienced nausea in the user testing reported, that VR should not be used in schools at all. For this reason, it is important to make the first and every experience as comfortable as possible. The anxiety related to digital devices which some students suffer [118], is as well a feature that needs to be taken into consideration and highlights the importance of comfort.

Through testing, it became clear that the nausea-causing elements and mechanisms are not yet properly recognized and fully understood in VR [250]. All actions to avoid the experience of motion sickness were made in the application design process, even though, the appearance of the phenomenon could not be prevented. The free movement (i.e., locomotion) was disabled in hand controllers and turning, and moving short distances was made possible only by physically turning around or moving. This was made to avoid conflict between senses of balance and sight since it is recognized as the cause of motion sickness [249]. The long-distance movement was enabled through a teleport option. In virtual worlds teleportation has been considered to be a better option to moving as free locomotion [250]. However, it can be argued whether the current version of teleportation is the best for people who suffer motion sickness or is it made for gamers willing to move even faster from one place to another.

The field of VR is fast evolving and after the user testing of this study, Apple (2024) has published new instructions for XR designers. To avoid anchored views, to create apps with the need for minimal movement, to avoid fast movements and to fade out content while it is moving and to face it after settled are recommended [235]. Meta also offers multiple options to customize the VR experience today, and for example, in Horizon Worlds [36] it is possible to narrow the field of view while moving (i.e., vignette effect). These features are something to consider adding in application and then retesting. It can be considered how much movement is actually needed to have in the learning material and to use movement only in places needed.

Another possibility to improve the developed application relates to the design of the environment. The use of smoother terrain or fewer mountains and hills minimizes the vertical movement and possibly increases the comfort level of students. Vertical movement in VR may cause similar symptoms for people as seasickness (cf. cyber sickness). Enabling users to teleport longer distances can reduce the number of actions needed and decrease the experience of discomfort. Of course, it can also be considered what to use VR for. If motion sickness cannot be fully removed, VR can, for example, in some level substitute school excursions [273] to different locations. The feature that causes students to feel nausea may not be seen as harmful while visiting an amusement park instead of experiencing nausea with regular learning.

With VR usage it must also be checked whether all safety features of the headset are used properly. For example, is the guardian on and working and are headsets adjusted (e.g., head stripes and pupil distance) in a comfortable way. Technical improvements, for example, lighter products must still be waited, but overall, a long way from the first very heavy versions [155] has been come. Z-fighting problems (i.e., two textures are fighting since overlapping with each other and having the same depth) [255] were fixed in the developed application, but it was not possible to remove the image vibration (i.e., a hot day or air thermal vibration effect). This was a problem, students pointed out. The phenomenon was caused by a compatibility problem of shaders and became visible after the built of application. To be more precise, it was a problem between the Terrain Tools plugin and standalone build. This element can possibly cause headache and migraine and relates to the R11. “It is safe to use” (medium priority defect). The importance of selecting functioning tools that do not conflict with the device’s properties must be taken carefully into account and to take time to choose the best tool options. In addition, it would help developers work if every tool had a list of supported devices.

The Terrain Tools plugin that was found problematic earlier, enabled building different landforms, to use the mixture of textures and to include realistic vegetation with animations (e.g., the grass swaying in the wind). The plugin was used to make the environment more real. Simultaneously, the environment was kept very simple with minimum objects to avoid cognitive overload [176] and to reduce the size of the application, which can slow the performance down. Three students suggested improving the environment and graphics. The shader problem is something to start with and to determine whether it is possible to fix the issue and continue using the plugin or replace it with something else in the upcoming versions.

The third challenge that has been found in Alalwan’s et al. (2020) study relates to the students’ lack of focused attention in learning, poor communication with the teacher, and even addictive behavior after using VR for longer periods of time. The addiction was a concern raised by parents participating in the Alalwan's study, but it strongly influenced teachers' opinion on the use of VR technology. It was revealed that Finnish teachers were not extremely concerned about the possibility of addiction related to VR learning, but they somewhat agreed on the statement 17.I. "The usage of the application doesn't expose students to game addiction" (M = 4.0). One teacher reported being more concerned about the topic of game addiction regarding games students play in their free time and in homes instead of material used in school lessons. However, addiction is a phenomenon that exists with digital games and is beneficial to consider on some level from design aspect [274]. Some students may be more susceptible to game addiction than others [275]. However, the problem can become visible from teachers' perspective as poor communication whatever the cause. In this study, teachers were hoping that students were not able to see each other or the teacher in the virtual environment. This was the least preferred feature in the teacher's questionnaire. It was a concern that if students would see each other more disturbance and chatting might be expected, and this could possibly lead to a situation with less learning and poorer communication with the teacher.

Overall, students reported preferring to use VR in schools significantly more often than teachers. From 5 to 10 times during the school year was the average of students answer (M = 2.7). Teachers were more moderate in their opinion thinking the frequency should be one or a couple of times during the school year (M = 3.5). Students may have a tendency to give rather positive than negative evaluations, and the novelty factor [118] was not similarly considered in this comparison as comparing the success of implementation of the requirements. It is notable, that students were more eager to use VR technology even with the shortcomings they experienced. The reason for that can be that they have faith that the technology will improve with time and issues that rose up can be solved which enhance the experience eventually.

7.2.4 The lack of VR learning materials

According to Alalwan et al. (2020) the fourth and the most common challenge that teachers are facing and half of them suffer is not finding suitable VR material for learning purposes [270] [213]. Useful and high-quality learning material was found important in this study as well. Overall, nine of twelve quality requirements discovered from teachers' responses related to the usefulness theme. To be more precise five of the statements connected to student's experience, two to usability, and one to safety and health theme as well. Two of the requirements, R1 and R3 measured the pure subject.

In previous studies focused on the effectiveness of DGBL, the results have been mixed [276] [277] [278]. It has been found that the quality of research design varies which impacts results. For example, if a quasi-experimental design is used instead of randomized trial, the learning discovered has been notably higher. [279] [280] According to Clark's et al. (2016) meta-analysis, learning outcomes are found significant in noncollaborative and noncompetitive games only which may explain why teachers in this study preferred that students were not able to see each other in VR environment. Multiple game play sessions and the average of six-hours playing are needed to gain significant difference and better learning outcomes than with nongame (i.e., other media) conditions. [280] In VR learning it is found that with over two-hour session the effectiveness suffers [24].

Even though the test setting of this study was not favouring learning because of a strict time frame, students agreed saying 5A. "The VR-game helped me to learn new species". However, only two students were able to name species correctly. According to feedback received collectively, three of the possible 18 species were remembered. Repetition that is recognized to be important for learning with games remained minimal [20]. Catching the same species multiple times while playing was not very likely since species appearing into the environment were randomly selected by the game. The testing time of each group was about 20 minutes instead of six hours [280]. In that time, students needed to familiarize themselves with the new device and the new environment, learn how to act, to find species and to learn those species. All these tasks together appeared to be too overwhelming to achieve in this time frame. Students reported difficulties especially with gripping and holding items in hands and collecting species.

It can be considered why most students reported learning species even though not learning them, and gave a more positive evaluation than the actual case was. Multiple possible reasons can be considered for that. Firstly, it may be possible that students actually perceived learning species but did not know exactly what was meant by the term species. Secondly, they may not have paid enough attention to what was asked since time shortage. Thirdly students may have recognized the importance of learning since being in school environment, were enjoying the use and saw the potential of the material even though the time limit challenged learning. On the other hand, this phenomenon may be also connected to the Makransky et al.'s (2019) findings of the enhancement of perceptions when using VR.

However, when single statements can be considered succeeded in the overall examination which utilizes the whole data collected, the inadequacies of the theme usefulness are recognized. The need to repair elements connected to quality requirements R1. "Material obeys the curriculum" (medium priority defect) and R3. "Material intensifies the knowledge of students" (high priority defect) are found which implies that analysis and calculations made are working well. To resolve the issue, can be started by adding time, and instructions, and see if the problem was related to these or will the material need other modifications.

The fact that teachers and students considered the usefulness of the learning material important is expected. Based on the definition of the concept, only teachers can decide whether the material can be used as learning material [98] and with *didacticized learning materials* such as educational games not only content is important but inscribed (i.e., invisible) knowledge of learning as well. [99] Teachers have the best knowledge of what the good learning material should include and in the best case scenario, they would be making the VR learning material themselves for their students. However, teachers and students are very dependent on material made by others, since there are only very few in general or expensive platforms and tools available to easily create VR content. On the other hand, making VR content requires time or an extra effort to learn to use new tools, which teachers possibly do not have.

It appears that people making material now lack required skills and teachers and students are not getting what is needed. Offering schools inadequate materials to use can impact their willingness to use the technology in the future too. Designing learning materials requires knowledge as has been observed in previous studies [21] [56]. Making valuable didacticized learning material the knowledge of the usefulness and usability aspect as well as pedagogy is required [99], and people making these materials should acquire themselves required knowledge. Interviewing professional teachers can help. However, as noted before, it is difficult to meet the required quality and to shift the knowledge of teaching professionals into the material.

In recent years, people have noticed how the field of education has turned into marketplace [80]. However, from a marketing aspect, it appears that education is not seen as interesting enough to invest a sufficient amount of time, money and effort [21] [56]. For some people making material for education can potentially be seen as an alternative source of income or perhaps a way to realize own personal dreams and fantasies of being a teacher. People are expecting a lot since people working with ICT have been considered being exceptionally talented [281]. However, Ljungqvist and Sonesson (2022) have noticed that ICT professionals have challenges in listening and making collaboration with teaching professionals [282] which can affect, for example, the quality of the digital learning materials designed. Tienari (1991) observed that IT-professionals are the first group that seems to be loyal not to the employer, but only their own profession and old patterns of behaviour may be persistent and still visible today. The need for professional code of conduct for ICT professionals is also seen important [281]. When modern people are living the Red Queen hypothesis true (i.e., the requirement of continuous running [4]), people making material for education may be still settling for less. Staying at the forefront of digitalization [1] [5] [6] requires constant work from all parties.

In designing learning material, instead of making subject specific material, a better goal could be to make functional learning materials, for example, tools and platforms that teachers and students can utilize by making material themselves. Implementing features that enable made material to meet all the requirements teachers and students are expecting is something to achieve. Seeking for standardized interface and paying extra attention and focusing on to usability and the ease of use aspect that are in many cases seen problematic and not forgetting the safety aspect either is vital, especially in educational context [160] [16] [192].

8 Conclusions

VR is found advantageous in the terms of learning in many ways, however, the research of VR learning is lacking and because of that the use of the technology is relatively random in schools which was discovered in this study as well. This case study provides insights and knowledge of the design of VR learning materials identified through the exploratory approach and experiments in an authentic school environment. The four phase structure of this study follows principles of the DSR and is a combination of the PST framework, the HCD process cycle and the LISD model.

The features of VR materials that teachers and students considered important were discovered through questionnaires, and user testing of a purpose-built VR application. 18 teachers and 17 students participated the study. The invitation to the online survey was published in a professional learning community (i.e., private teacher’s Facebook group). A class of seventh grade students who studied in a science-oriented line conducted user testing. Testing was performed in groups of three to four students during two lessons as part of normal classroom work. Students tested a VR application that contained features that were identified as very important by the teachers of the study. Testing was conducted with Meta Quest 2 headsets. The collected data were processed and statistically analysed to answer the research questions.

- RQ 1. What are the main components of valuable VR learning material?
- RQ 2. What kind of role are teachers preferred to have during the VR session?
- RQ 3. Does teaching experience have an impact on preferred VR features?
- RQ 4. Is there a difference between teacher’s and student’s VR preferences?
- RQ 5. How can VR learning be implemented into education?
- RQ 6. What are the challenges of VR learning?

RQ 1. The study revealed twelve quality requirements which are essential to implement into VR learning materials. Over 80% of the requirements related to the themes of usefulness and student’s experience, but the usability as well as safety and health aspects were also found very important. Usefulness is a theme that teachers are expecting and is shown in previous studies to be an attribute they struggle to find. RQ 3. The importance of all themes studied is evaluated higher by more experienced teachers, which implies that they are ready to require better learning materials for students whereas teachers with less teaching experience are willing to accept materials with lower criteria. RQ 2. The study revealed that teachers do not have a strong opinion about the role they would prefer to have during a VR session, but the passive role was slightly more favored. Teachers with over ten years of teaching experience favor significantly more the usability aspect of the material, and having a more passive role for themselves in VR learning. This can be achieved in VR applications, for example, with more comprehensive instructions, which is a feature that is also favored by students.

RQ 4. Overall, the themes teachers considered important are very similar to students' perceptions. However, there was a significant difference in terms of frequency and the students hoped that VR lessons would be held more often than the teachers preferred. Offering a positive learning experience for students is important for teachers. This is also a challenge as it has been found in previous study that students' interest and anxiety in digital devices varies. In this study, the implementation of the theme of student's experience was succeeded and students reported liking the game and having fun.

RQ 5. Similar to previous studies, the results relating to the overall learning conditions of VR revealed that biology teachers considered VR learning to have a potential and described the conditions for VR learning in education. They were able to name many topics such as ecosystems that could be learned with VR and students were able to see even wider variety of uses in schools. Teachers were thinking that a suitable length for a VR session would be over ten but not more than 40 minutes, and there should be a VR headset if not for every student, then at least for every other student in the classroom. Having many headsets in a classroom simultaneously can require special arrangements and is something that needs to be taken into account in a design perspective to allow sufficient movement for students.

RQ 6. In testing, it was found that the core needs of teachers and students were met with the built-in application on many levels, but more work is still needed. The usefulness and the safety aspects are especially something to improve and to focus on in the future studies. It is important to enable test settings where subject related learning in an authentic classroom environment can also be studied. Nausea and elements that can cause headache and migraine must be carefully taken into account and harmful impacts minimized. A disadvantage teachers reported was the relatively short life cycle of digital devices. This is seen problematic among teachers because, after implementing VR learning into education, it will require further investments, thus meaning VR is competing with other acquisitions that may be more critical. The cost of devices is found a challenge in previous studies as well. Instructions are something to pay particular attention to, because the users' freedom of choice allows to choose where to look in the VR environment. It is possible to accidentally ignore the instructions and get in trouble.

It is proposed to design primarily functional learning materials (i.e., tools and platforms) that would allow teachers and students to make the learning materials themselves, meet the diverse needs, and increase usability. The participatory design approach revealed its value, and as a whole it was discovered that the combination of design frameworks chosen appeared to be successful in terms of identifying and testing the features to implement into VR learning. Teachers' preferences were recognized through a questionnaire. Furthermore, the custom application facilitated the study of VR learning and the features important to include also from the students' perspective. If the triangular of the PST is considered, the results show that limitations in the implementation of all three areas (pedagogy, space and technology) were found. A useful application for testing VR learning was developed in the study based on the results of the classroom experiment. Using a similar design approach for designing VR learning materials and studying VR learning can be recommended for future studies.

The five main findings of the study are:

1. Follow the principles of participatory design and involve stakeholders in the learning material design process.
2. Utilize research findings and select the type of learning material (e.g., functional learning material) according to the available knowhow.
3. The usefulness of the material and the student's experience perspectives must be taken into account in particular.
4. Pay special attention to the instructions. Take advantage of usability heuristics and standards to ensure the ease of use. Perform user tests in an authentic classroom environment and with adequate sample size.
5. Identify the safety and health challenges in order to avoid elements in the learning materials that can cause, for example, discomfort and nausea.

Acknowledgments

I would like to thank my instructors Teijo Lehtonen and Taneli Nyyssönen for their expertise and patient guidance. From the department of teacher education, I would like to thank Marjaana Veermans for all her aid and expertise, especially in relation to the chapter on education and Eero Laakkonen for his expertise in data analysis and in data processing. I would also like to thank to Sanna Virtanen for her expertise and help with the English language.

References

[1] E. Comission, "Digital Decade 2025, Country Reports, Finland," Brussels, 2025.

[2] E. Commission, "Digital Decade DESI visualisation tool," European Commission, 2025.

[3] J. Fagerlund, K. Leino, M. Niilo-Rämä, E. Puhakka and I. Markkanen, "Kohti digiosaamisen strategista kehittämistä, Kansainvälinen monilukutaidon ja ohjemoinnillisen ajattelun tutkimus (ICILS 2023)," Koulutuksen tutkimuslaitos, Jyväskylä, 2024.

[4] L. Carroll, Through the Looking-Glass And What Alice Found There, The Millennium Fulcrum Edition 1.7, 1865.

[5] L. V. Valen, "Two Models of evolution," *Nature,* vol. 252, no. 5481, pp. 298-300, 1974.

[6] L. Mûller, "Chapter Thirty Six - 1973 The red queen hypothesis," in *Conseptual Breakthroughs in Evolutionary Ecology*, Academic press, 2020, pp. 85-86.

[7] M. Morand, "ASCM Insights, The Red Queen Hypothesis," 5 19 2025. [Online]. Available: https://www.ascm.org/ascm-insights/the-red-queen-hypothesis. [Accessed 10 10 2025].

[8] R. Linturi and O. Kuusi, "Suomen sata uutta mahdollisuutta 2018-2037," Eduskunnan tulevaisuusvaliokunta 1/2018, Helsinki, 2018.

[9] P. Banarjee, "A Review of: "Virtual Reality Technology, 2nd edition" Burdea, Grigore ; Coiffet, Philippe," *IIE Transactions,* no. Vol.37 (7), p. 682, 2005.

[10] I. E. Sutherland, "The Ultimate Display," *Information Processing Techniques, Office ARPA, OSD,* 1965.

[11] M. Brown, M. McCormack, J. Reeves, D. C. Brooks and S. Grajek, "2020 EDUCAUSE Horizon Report™ | Teaching and Learning Edition," *Horizon Report,* 2020.

[12] C. Grimsdale, "Virtual reality evolution or revolution," *IEE Colloquium on Distributed Virtual Reality,* pp. 2/1-2/4, 1993.

[13] S. LaValle, Virtual Reality, Cambridge University Press, 2020, p. 7.

[14] T. I. o. E. a. E. Engineers, IEEE Standars Glossary of Software Engineering Terminology., New York: IEEE Standard 610.12-1990, 1990.

[15] ISO, "ISO 9241-11:2018," ISO, 2 2018. [Online]. Available: https://www.iso.org/standard/63500.html. [Accessed 22 7 2025].

[16] D. Norman, The Design of Everyday Things Revised and Expanded Edition, New York: Basic Books, 2013.

[17] D. Schuler and A. Namioka, Participatory design, Principles and Practices, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1993.

[18] A. Parmaxi, "Virtual reality in language learning: a systematic review and implications for research and practice, Interactive Learning Environments," vol. 31, no. 1, pp. 172-184, 2023.

[19] D. Panchuk, M. Klusemann and S. Hadlow, "Exploring the Effectiveness of Immersive Video for Training Decision-Making Capability in Elite, Youth Basketball Players," *Front Psychol.,* vol. 9, no. 2315, 2018.

[20] C. González-Gonzáles and F. Blanco-Izquierdo, "Designing social videogames for educational uses," *Computers & Education,* vol. 58, pp. 250-262, 2012.

[21] A. Ibañez-Etxeberria, C. Gómez-Carrasco, O. Fontal and S. García-Ceballos, "Virtual Environments and Augmented Reality Applied to Heritage Education. An Evaluative Study. Applied Sciences.," vol. 10(7):2352, 2020.

[22] B. Kyaw, N. Saxena, P. Posadzki, J. Vseteckova, C. Nikolaou, P. George, U. Divakar, I. Masiello, A. Kononowicz, N. Zary and C. Tudor, "Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration.," *J Med Internet Res.,* vol. 22;21(1):e12959, 2019.

[23] E. Eva Eriksson, C. Heath and P. P. P. Ljungstrand, "Makerspace in school—Considerations from a large-scale national testbed," *International Journal of Child-Computer Interaction,* vol. 16, pp. 9-15, 2018.

[24] R. Villena-Taranilla, S. Tirado-Olivares and R. Cózar-Gutiérrez, "Effects of virtual reality learning outcomes in K-6 education: A meta-analysis," *Educational Research Review,* vol. 35, 2022.

[25] P. R. de Lope, N. Medina-Medina, P. Paderewski-Rodríguez and F. L. Vela, "Design methodology for educational games based on interactive screenplays," *CEUR Workshop Proceedings,* 2015.

[26] M. Zyda, "From Visual Simulation to Virtual Reality to Games," *Perspectives, IEEE Computer Society,* 2005.

[27] T. M. Connolly, E. A. Boyle, E. H. T. MacArthur and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Computers & Education,* vol. 59, no. 2, p. 661, 2012.

[28] H. Desurvire, M. Caplan and J. Toth, "Using Heuristics to evaluate the Playability of Games," in *Human Computer Interactions ACM, 24-29th April*, Vienna, 2004.

[29] M. Brown, "Evaluating Computer Game Usability: Developing Heuristics Based on User Experience," *ResearchGate,* pp. 1-6, 2008.

[30] M. Lui and J. D. Slotta, "Immersive simulations for smart classrooms: exploring evolutionary concepts in secondary science," *Technology, Pedagogy and Education,* vol. 23, no. 1, pp. 57-80, 2014.

[31] Meta, "Meta Quest Store," Meta, [Online]. Available: https://www.meta.com/en-gb/experiences. [Accessed 12 10 2025b].

[32] Meta, "Meta Horizon Workrooms," Meta, [Online]. Available: https://forwork.meta.com/fi/horizon-workrooms. [Accessed 4 5 2025].

[33] Meta, "Meta Horizon," Meta, [Online]. Available: https://horizon.meta.com. [Accessed 14 7 2025b].

[34] P. Rospigliosi, "Metaverse or Simulacra? Rpblox, Minecraft, Meta and the turn to virtual reality for education, socialisation and work," *Interactive Learning Environments,* vol. 30, no. 1, pp. 1-3, 2022.

[35] T. Tamminen, "EduFinland – Finnish education in Second Life," 2014. [Online]. Available: https://eoppimiskeskus.fi/projekti/edufinland-fi-finnish-education-in-second-life/. [Accessed 11 28 2023].

[36] Meta, "Meta Horizon Worlds," Meta, [Online]. Available: https://www.meta.com/en-gb/experiences/worlds/2532035600194083/. [Accessed 4 5 2025].

[37] E. VR, "VR education: Transforming Learning Through Immersive Experience like Titanic VR," Engage, [Online]. Available: https://engagevr.io/. [Accessed 4 5 2025].

[38] E. xr, "Library," Eon Xr, [Online]. Available: https://core.eon-xr.com/Library/Index_V2. [Accessed 4 5 2025].

[39] S. life, "Explore discover create," Second life, [Online]. Available: https://secondlife.com/. [Accessed 4 5 2025].

[40] VrChat, "Create a Universe of Endless Possibilities," VrChat, [Online]. Available: https://hello.vrchat.com. [Accessed 4 5 2025].

[41] Zapworks, "Zappworks studio," Zappar, [Online]. Available: https://zap.works/studio/. [Accessed 4 5 2025].

[42] Ctrl, "Ctrl.studio," Ctrl, [Online]. Available: https://ctrl.studio/. [Accessed 4 5 2025].

[43] Wonda, "Train Smarter, Not Harder," Wonda, [Online]. Available: https://www.wonda,pro/. [Accessed 4 5 2025].

[44] U. b. ViewSonic, "Your Virtual Campus for Online Learning," Universe by ViewSonic, [Online]. Available: https://universe.viewsonic.io/. [Accessed 4 5 2025].

[45] Eduverse, "Introducing Eduverse... The World's First K-12 Metaverse," Eduverse, [Online]. Available: https://eduverse.com. [Accessed 4 5 2025].

[46] Roblox, "Roblox," Roblox, [Online]. Available: https://www.roblox.com. [Accessed 4 5 2025].

[47] O. Chrast and J. Beseda, "New Technologies and New ICT and Gamification-Based Approaches for the Cultural Heritage Education and Comparison of their use in some Czech Republic's Cultural Heritage Sites, Cultural Monuments and Virtual Museums," *Proceedings of the Overcoming the Challenges and Barriers in Open Education,* p. 401–411, 2018.

[48] J. García-Fernández and L. Mateus, "Solution supporting the communication of the built heritage: Semi-automatic production path to transfer semantic LIDAR data to minecraft environment," *Digital Applications in Archaeology and Cultural Heritage,* vol. 14, 2019.

[49] C.-C. Lin, "Challenging the future: Development of students' multimodal competencies by designing an online exhibition," *Entertainment Computing,* vol. 48, 2024.

[50] S. Pedersen, P. Malmberg, A. J. Christensen, M. Pedersen, S. Nipper, C. D. Graem and e. al., "E-learning Nordic 2006. Impact of ICT on education," 2006.

[51] L. Ilomäki and M. Lakkala, "Koulu, digitaalinen teknologia ja toimivat käytännöt," *Opetusteknologia koulun arjessa II,* 2011.

[52] J. Kaisto, T. Hämäläinen and S. Järvelä, "Tieto- ja viestintätekniikan pedagoginen vaikuttavuus pohjoisessa Suomessa," *University of Oulu,* 2007.

[53] R. Deaney, K. Ruthven and S. Hennessy, "Pupil perspectives on the contribution of information and communication technology to teaching and learning in the secondary school," *Research Papers in Education,* vol. 18, pp. 141-165, 2003.

[54] L. Ilomäki, T. Jaakkola, L. N. S. Nirhamo, E. Lehtinen, T. Toikkanen, A. Kantosalo, S. Paavola, M. Lakkala, M. Veermans, A. Tapola and H. Muukkonen, "Laatua e-oppimateriaaleihin - E-oppimateriaalit opetuksessa ja oppimisessa," *Opetushallitus, Oppaat ja käiskirjat,* vol. 5, 2012.

[55] J. P. Gee, "Learning by Design: Good Video Games as Learning Machines," *E-Learning and Digital Media,* vol. 2, no. 1, pp. 5-16, 2005.

[56] A. All, E. P. N. Castellar and J. Van Looy, "Assessing the effectiveness of digital game-based learning: Best practices," *Computers & Education,* vol. 92–93, 2016.

[57] PennState, "Meta Quest 2 Experiences," Penn State Teaching and Learning with Technology, 11 1 2024. [Online]. Available: https://imex.psu.edu/experience-catalog/vr-experiences-retired/meta-quest-experiences/. [Accessed 29 11 2023].

[58] P. C. Cozby and S. C. Bates, Methods in Behavioral Research, New York: McGraw-Hill Education, 2015.

[59] M. Sein, O. Henfridsson, S. Purao, M. Rossi and R. Lindgren, "Action Design Research," *MIS Quarterly,* vol. 35, no. 1, pp. 37, 38, 40, 2011.

[60] J. G. Walls, G. R. Widmeyer and O. A. El Sawy, "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research,* vol. 3, no. 1, pp. 36-37, 1992.

[61] M. Markus, A. Majchrzak and L. Gasser, "A Design Theory for Systems That Support Emergent Knowledge Processes," *MIS Quarterly,* vol. 26, pp. 179-212, 2002.

[62] A. R. Hevner, S. T. March, J. Park and S. Ram, "Design Science in Information System Research," *Mis Quarterly,* vol. 28, no. 1, pp. 82-90, 2004.

[63] J. vom Brocke, A. Hevner and A. Maedche, "Introduction to Design Science Research," *Design Science Research, Cases, Cham,* 2020.

[64] Kouluhallitus, Peruskoulun opetussuunnitelman perusteet 1985, Kouluhallitus : Valtion painatuskeskus 1985, 1985.

[65] P. Saarikoski, "Koneen ja koulun ensikohtaaminen: suomalaisen atk-koulutuksen varhaisvaiheet peruskoulussa ja lukiossa.," *Tekniikan Vaiheita,* vol. 24, no. 3, pp. 5-19, 2006.

[66] S. Rahikainen, "Tietotekniikan murros Suomen peruskouluissa 1985-1986 Opettaja-lehtien tekstien kuvaamana," Jyväskylän yliopisto, 2016.

[67] E. Tanhua-Piironen, K. S-S, M.-T. Kaarakainen and J. Viteli, Digiajan koulu II, Opetus ja kulttuuriministeriön julkaisuja 2020:17, Helsinki: Opetus ja kulttuuriministeriö, 2020.

[68] R. Tamim, R. Bernard, E. Borokhovski, P. Abrami and R. Schmid, "What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study," *Review of Educational Research,* vol. 81, no. 1, pp. 4-28, 2011.

[69] S.-S. Kaarakainen and M.-T. Kaarakainen, "Tulevaisuuden kansalaisia rakentamassa: Uudet lukutaidot koulutuksen ja opetuksen digitalisaation kehityksessä," in *Uusia lukutaitoja rakentamassa - Building new literacies. AFinLA vuosikirja 2018*, Jyväskylä, Suomen soveltavan kielitieteen yhdistys AFinLA, 2018, pp. 22-40.

[70] P. Siljander, Systemaattinen johdatus kasvatustieteeseen, Keuruu: Otava, 2005, pp. 51-53, 75-76.

[71] Valtioneuvosto, "Valtioneuvoston selonteko: Suomen digitaalinen kompassi," Valtioneuvoston julkaisuja 2022:65, Helsinki, 2022.

[72] S. Oinas and M.-P. Vainikainen, Digitalisaatio oppimisen ja oppimistulosten selittäjänä, Jyväskylä: Jyväskylän yliopistopaino, 2024.

[73] T. Fawns, "An entangled pedagogy: Looking beyond the pedagogy - Technology dichotomy," *Postdigital Science and Education,* vol. 4, no. 3, pp. 711-728, 2022.

[74] J. Voogt and N. Roblin, "A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies," *Journals of Curriculum Studies,* vol. 44, no. 3, pp. 299-321, 2012.

[75] J. Reis, M. Amorium, N. Meläo, Y. Cohen and M. Rodrigues, "Digitalization: A Literature Review and Research Agenda," *Springer Nature,* pp. 443-456, 2020.

[76] F. Pettersson, "Understanding digitalization and educational change in school by means of activity theory and the levels of learning concept," *Education and Information Technologies,* vol. 26, no. 1, pp. 187-204, 2021.

[77] L. Savolainen, Who is the algorithms? Interfacing the social, emotional, and algorithmic, Helsinki: Unigrafia, 2023.

[78] K. Leino, E. Puhakka and M. Niilo-Rämä, "Tieto- ja viestintäteknologia koulujen arjessa - ICILS Opettajapaneeli 2020 -tutkimuksen tuloksia.," IEA, 2021.

[79] A. D. Olofsson, O. J. Lindberg and G. Fransson, "Students' Voices about Information and Communication Technology in Upper Secondary Schools," *International Journal of Information and Learning Technology,* vol. 35, no. 2, pp. 82-92, 2018.

[80] F. Peruzzo, J. Stephen and E. Grimaldi, "Peopling the crowded education state: Heterachical spaces, EdTech markets and new models of govering during the COVID-19 pandemic," *International Journal of Educational Research,* vol. 114, 2022.

[81] Opetushallitus, "Tietoturva ja -suoja koulussa," Opetushallitus, [Online]. Available: https://www.oph.fi/fi/koulutus-ja-tutkinnot/tietoturva-ja-suoja-koulussa. [Accessed 5 11 2025b].

[82] M. Muhonen, M.-T. Kaarakainen and J. Savela, "Opettajien teknologiataidot oppilaiden tulevaisuuden taitojen (epä)tasa-arvoisuuden edistäjänä?," in *Tuovi 13: Interaktiivinen tekniikka koulutuksessa 2015-konferenssin tutkijatapaamisen artikkelit*, Tampere, Tampereen yliopisto, 2015, pp. 56-64.

[83] R. Ahtiainen, M. Asikainen, L. Heikonen, N. H. R. Hienonen, P. Lindfors, L. E, M. Lintuvuori, S. Oinas, A. Rimpelä and M.-P. Vainikainen, "Koulunkäynti, opetus ja hyvinvointi kouluyhteisössä koronapandemian aikana : Ensitulokset," University of Helsinki, University of Tampere, Helsinki, 2020.

[84] L. Ilomäki, The effects of ICT on school: teachers' and students' perspectives, Turku: From the Department of Techer Education, University of Turku, 2008.

[85] R. Vehviläinen, "Tietotekniikan kansallinen järjestötoiminta, Tietotekniikan liitto - monipuolinen vaikuttaja," in *Tietotekniikan alkuvuodet Suomessa*, Jyväskylä, Gummeruksen Kirjapaino Oy, 1993, p. 439.

[86] Opetushallitus, Perusopetuksen opetussuunnitelman perusteet 2014, Määräykset ja ohjeet 2014:96, Helsinki: Opetushallitus, 2014.

[87] N. Ouakrim-Soivio, "Laaja-alaisen osaamisen tavoitteet ja niiden arviointi osana oppiaineen arviointia," in *Laaja-alainen osaaminen koulussa: Ajattelijana ja oppijana kehittyminen*, Helsinki, Gaudeamus, 2022, pp. 218-229.

[88] Opetushallitus, "Digitaalisen osaamisen kuvaukset," Opetushallitus, 2024. [Online]. [Accessed 3 12 2024].

[89] J. R. A. Carter, M. Rice, S. Yang and H. A. Jackson, "Self-regulated learning in online learning environments: Strategies for remote learning," *Information and Learning Sciences,* vol. 121, no. 56, p. 321–329, 2020.

[90] F. Sahlström, M. Tanner and C. Olin-Scheller, "Smartphones in classrooms: reading, writing and talking rapidly changing educational spaces.," *Learning, Culture and Social Interaction,* vol. 22, pp. 1-5, 2019.

[91] B. See, S. Gorard, B. Lu, L. Dong and N. Siddiqui, "Is technology always helpful? A critical review of the impact on learning outcomes of education technology in supporting formative assesment in schools," *Research Papers in Education,* vol. 37, no. 6, pp. 1064-1096, 2022.

[92] M. Vedechkina and F. Borgonovi, "A Rewiew of Evidence on the Role of Digital Technology in Shaping Attention and Cognitive Control in Children," *Frontiers of Psychology,* vol. 12, no. 611155, pp. 1-16, 2021.

[93] Opetushallitus, Perusopetuksen opetussuunnitelman perusteet 2014, Opetushallitus, 2014, pp. 694-699, 929-945.

[94] M.-T. Kaarikainen, O. Kivinen and V. T, "Performance-based testing for ICT skills assessing: A case study of students and teachers' ICT skills in Finnish schools," *Universal Access in the Information Society,* vol. 17, no. 2, pp. 349-360, 2018.

[95] M.-T. Kaarakainen and L. Saikkonen, "Remark on digital accessibility: educational disperities define digital inclusion from adolescence onwards," *Universal Access in the Information Society,* vol. 22, pp. 1279-1292, 2022.

[96] N. Autere, A. Talvinen, M. Hakala and J. Pöysä-Tarhonen, "Esteettömyys digitaalisissa ympäristöissä kielellisten erityisvaikeuksien näkökulmasta.," *Oppimisen ja oppimisvaikeuksien erityislehti NMI Bulletin Niilo Mäki instituutti,* vol. 29, no. 2, pp. 23-37, 2019.

[97] R. Jeffrey and D. Chisnell, Handbook of Usability Testing : How to Plan, Design and Conduct Effective Tests, 2nd ed, Indianapolis: John Wiley & Sons, 2008.

[98] P. A. Barker and L. M. Campbell, "Metadata for learning materials: an overview of existing standards and current developments," *Technology, Instruction, Cognition and Learning,* vol. 7, no. 3-4, pp. 225-243, 2010.

[99] J. Bundsgaard and T. I. Hansen, "EVALUATION OF LEARNING MATERIALS: A HOLISTIC FRAMEWORK," *Journal of Learning Design,* vol. 4, no. 4, pp. 31-42, 2011.

[100] A. All, E. P. N. Castellar and J. Van Looy, "Assessing the effectiveness of digital game-based learning: Best practices.," *Computers & Education,* Vols. 92-93, pp. 90-103, 2016.

[101] B. Haßler, L. Major and S. Hennessy, "Tablet use in schools: a critical review of the evidence for learning outcomes," *Journal of Computer Assisted Learning,* vol. 32, p. 139–156, 2016.

[102] Opetushallitus, "E-oppimateriaalin laatukriteerit," Opetushallitus, [Online]. Available: https://www.oph.fi/fi/julkaisut/e-oppimateriaalin-laatukriteerit. [Accessed 27 12 2023].

[103] J. F. Herbart, Pestalozzi's Idee eines ABC der Anschauund als ein Cyklus von Vorübungen im Auffassen der Gestalten, Göttingen, 1804.

[104] A. M. Durik and J. M. Harackiewicz, "Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. Journal of Educational Psychology,," vol. 99, no. 3, p. 597–610, 2007.

[105] K. A. Renninger, "Chapter 13 - Individual interest and its implications for understanding intrinsic motivation,," in *In Educational Psychology, Intrinsic and Extrinsic Motivation*, Academic Press, 2000, pp. 373-404.

[106] U. Schiefele, "Situational and individual interest," in *In K. R. Wentzel & A. Wigfield (Eds.) Handbook of motivation at school* , New York, NY: Routledge, 2009, pp. 197-223.

[107] M. Mitchell, "Situational interest: Its multifaceted structure in the secondary school mathematics classroom.," *Journal of Educational Psychology,* vol. 85, no. 3, pp. 424-436, 1993.

[108] S. Hidi and W. Baird, "Strategies for increasing text-based interest and students' recall of expository texts," *Reading Research Quarterly,* vol. 23, no. 4, p. 465–483, 1988.

[109] K. A. Renninger, "Children's play interests, representation, and activity," in *Knowing and remembering in young children*, Cambridge University Press., 1990, pp. 127-165.

[110] U. Schiefele and M. Csikszentmihalyi, "Motivation and Ability as Factors in Mathematics Experience and Achievement," *Journal for Research in Mathematics Education,* vol. 26, no. 2, p. 163–181, 1995.

[111] M. Csikszentmihalyi, K. Rathude and S. Whalen, Talented teenagers. The roots of success and failure, New York: Cambridge University Press, 1993.

[112] A. Krappa and M. Prenzel, "Research on Interest in Science: Theories, methods, and findings," *International Journal of Science Education* , vol. 33, no. 1, pp. 27-50, 2011.

[113] S. T. Coleridge, "Chapter XIV," in *Bibliografia Literaria*, The Project Gutenberg eBook, 1817.

[114] M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience.," 1990.

[115] T. Gunning, "Chapter Title: Attractions: How They Came into the World," in *The Cinema of Attractions Reloaded*, Amsterdam University Press, 2006.

[116] L. Ermi and F. Mäyrä, ". (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion..," *Worlds in Play: Int. Perspectives on Digital Games Research.,* 2005.

[117] S. Koivuhovi and K.-M. Polso, "Oppilaiden motivaatio ja suhtautuminen digivälitteiseen oppimiseen," in *Digitalisaation vaikutus oppimiseen, oppimistilanteisiin ja oppimistuloksiin. DigiVOO-hankkeen väliraportti 2022.*, Tampere University and University of Helsinki, 2022, pp. 78-92.

[118] R. Hotulainen and S. Oinas, "Itsearvioidut digitaalisen oppimisen taidot," in *Digitalisaation vaikutus oppimiseen, oppimistilanteisiin ja oppimistuloksiin. DigiVOO-hankkeen väliraportti 2022*, Tampere University and University of Helsinki, 2022, pp. 59-68.

[119] D. T. Brookes, "Social positionong in small group interactions in an investigative science learning environment physics class," *Physics review physics education research,* vol. 17, no. 1, 2021.

[120] E. Etkina, "Millikan award lecture: Students of physics—Listeners, observers, or collaborative participants in physics scientific practices?," *American Journal of Physics ,* vol. 83, no. 8, pp. 669-679, 2015.

[121] K. Hakkarainen and P. Seitamaa-Hakkarainen, "Learning by Inventing - Theoretical Foundations," in *Invention Pedagogy – The Finnish Approach to Maker Education*, University of Helsinki, 2022, pp. 15 - 27.

[122] S. Riikonen, P. Seitamaa-Hakkarainen and K. Hakkarainen, "Bringing maker practices to school : tracing discursive and materially mediated aspects of student teams' collaborative making processes," *International Journal of Computer-Supported Collaborative Learning,* vol. 15, no. 3, pp. 319-349, 2020.

[123] E. Nuutila, S. Törmä and L. Malmi, "PBL and Computer Programming — The Seven Steps Method with Adaptations," *Computer Science Education,* vol. 15, no. 2, pp. 123 - 142, 2005.

[124] A. Jevremovic, G. Shimic, M. Veinović and N. Ristić, "IP Addressing: Problem-Based Learning Approach on Computer Networks," *IEEE Transactions on Learning Technologies,* vol. 10, no. 3, p. 368, 2017.

[125] H. G. Schmidt, "Problem-based learning: rationale and description," *Medical Education,* vol. 17, no. 1, pp. 11-16, 1983.

[126] P. Mattila and P. (. Silander, How to create the school of the future - Revolutionary thinking and design from Finland, Oulu: University of Oulu, 2015.

[127] W. D. Pérez, D. L. Fields and D. March, "Innovations and Challenges: Conceptualizing CLIL Practice," *Theory Into Practice,* vol. 57, no. 3, pp. 177-284, 2018.

[128] D. R. Towns-Belton, "Teacher'sReported Use of Phenomenon-Based Learning in Second STEM Classrooms," The School of Graduate and Professional Studies, Department of Teaching and Learning, 2022.

[129] S. Paavola and K. Hakkarainen, "The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning," *Education Sciences,* vol. 14, p. 535–557, 2005.

[130] K. Hakkarainen and S. Paavola, "Towards Trialogical approach to learn," Academia, 2013.

[131] N. Sansone, D. Cesareni, I. Bortolotti and K. McLay, "The Designing and Re-Designing of a Blended University Course Based on the Trialogical Learning Approach," *Education Sciences,* vol. 11, no. 951, 2021.

[132] D. Radcliffe, H. Wilson, D. Powell and B. Tibbetts, "Designing Next Generation Places of Learning: Collaboration," The University of Queensland, 2008.

[133] JISC, Designing Spaces for Effective Learning: A Guide to 21st Century Learning Space Design, Joint Information Systems Committee, 2006.

[134] D. Radcliffe, "A Pedagogy-Space-Technology (PST) Framework for Designing and Evaluating Learning Places," *Proceedings of the Next Generation Learning Spaces 2008 Colloquium,* pp. 11-15, 2009.

[135] J. Slotta, "Evolving the classrooms of the future: The interplay of pedagogy, technology and community," *Classroom of the Future, Leiden, The Netherlands: Brill,* 2010.

[136] B. Reiser, I. Tabak, W. Sandoval, B. Smith, F. McFarland and A. Leone, "BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms," 2001.

[137] Plato, The Allegory of the Cave, 380 BCE.

[138] R. Descartes, Principia philosophiae, 1644.

[139] G. Burdea, P. Coiffet and P. Banerjee, "A Review of: "Virtual Reality Technology, 2nd edition," *IIE Transactions,,* no. Vol.37 (7), pp. p.682-683, 2003.

[140] N. S. Association, *Stereo World,* vol. 16, no. 3, 1989.

[141] I. Kant and S. T. Soemmerring, Über das Organ der Seele, Köningsberge, 1796.

[142] S. Bryson, "Virtual Reality: A Definition History - A Personal Essay," *arXiv.org,* pp. 1-6, 2013.

[143] G. Burdea and P. Goiffet, "La Réalité Virtuelle," *Hermes, Paris,* 1993.

[144] S. Benford, C. Brown, G. Reynard and C. Greenhalgh, "Shared Spaces: Transportation, Artificiality, and Spatiality," in *CSCW '96 Proceedings of the 1996 CM*, Nottingham, 1996.

[145] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE TRANSACTIONS on Information and Systems,* Vols. E77-D, no. 12, pp. 1321-1329 , 1994.

[146] C. Arth, R. Grasset, L. Gruber, T. Langlotz, A. Mulloni and D. Wagner, "The History of Mobile Augmented Reality," *Computer graphics and vision, arXiv,* 2015.

[147] M. Speicher, B. Hall and M. Nebeling, "What is Mixed Reality?," *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery,,* pp. 1-15, 2019.

[148] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, H. T. Luan and X. Shen, "A Survey on Metaverse: Fundamentals, security, and Privacy," *IEEE Communications Survey & Tutorials,* vol. 25, no. 1, 2023.

[149] T. Immonen, "Metaverse-Based Demonstrators as an Alternative to Traditional Presentations: Case Fossil-Free Steelmaking Processes," University of Turku, Turku, 2025.

[150] Sensorama, *Introducing... sensorama. The Revolutional Motion Picture System that takes you into another world.,* 1960.

[151] *Morton Heilig's Sensorama (Interview).mov.* [Film]. 2011.

[152] *The Sensorama - The Precursor to Virtual Reality Technoliogy.* [Film]. 2021.

[153] M. Heilig, "Sensorama". USA Patent 3,050,870, 28 8 1962.

[154] M. Heilig, "Stereoscopic-television apparatus for individual use". USA Patent 2,955,156, 4 10 1960.

[155] I. E. Sutherland, "A head-mounted three dimensional display. Seminal graphics: pioneering efforts that shaped the field.," *Association for Computing Machinery,* p. 295–302, 1998.

[156] J. J. Batter and J. F. P. Brooks, "GROPE-1: a computer display to the sense of feel. Netherlands. [not original found]," pp. 759-763, 1971.

[157] C. Blanchard, S. Burgess, Y. Halvill, J. Lanier, A. Lasko, M. Oberman and M. Teitel, "Reality Built For Two : A Virtual Reality Tool," *VPL Research, Inc,* 1990.

[158] C. Cruz-Neira, D. J. Sandin and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," *In Proceedings of the 20th annual conference on Computer graphics and interactive techniques ,* 1993.

[159] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon and J. C. Hart, "The CAVE: audio visual experience automatic virtual environment.," *Communications of the ACM,* vol. 35(6), 1992.

[160] B. Shneiderman, Designing the User Interface, United States of America: Addison-Wesley, 1992, pp. 2, 10, 11, 21-31, 66-69, 72-73, 222-228, 472-493, 495-500, 506-530, 537.

[161] Pico, "Pico Motion Tracker," Pico, [Online]. Available: https://www.picoxr.com/global/products/pico-motion-tracker. [Accessed 17 11 2025b].

[162] Meta, "Meta Quest 2, Tech specs".

[163] J. Donovan, "Market Overview and Market Forecasts for the VR Business," *Proceedings of Virtual Reality Systems '93 Conference, SIG-Advanced Applications,* pp. 25-28, 1993.

[164] C. Machover, " IEEE Computer Graphics and Applications," *The business of computer graphics ,* vol. 20, no. 1, p. 45, 2000.

[165] Google, "Lataa Google Cardboard -sovellus," Google, [Online]. Available: https://arvr.google.com/cardboard. [Accessed 18 8 2024].

[166] Samsung, "Gear VR with Controller (SM-R325)," Samsung, [Online]. Available: https://www.samsung.com/fi/mobile-accessories/gear-vr-r325-sm-r325nzvcnee. [Accessed 18 8 2024].

[167] Microsoft, "Xbox 360's "Project Natal" Awakens Imagination of Global Video-Game Industry," 24 9 2009. [Online]. Available: https://news.microsoft.com/source/2009/09/23/xbox-360s-project-natal-awakens-imagination-of-global-video-game-industry. [Accessed 13 11 2025].

[168] Vive, "Vive," HTC Vive, [Online]. Available: https://www.vive.com/eu. [Accessed 13 11 2025].

[169] S. LaValle, A. Yershova, M. Katsev and M. Antonov, "Head Tracking for Oculus Rift," *IEEE International Conference on Robotics & Automation (ICRA),* 2014.

[170] Meta, "Facebook to Acquire Oculus," 2014.

[171] PlayStation.Blog, "Introducing Project Morpheus," Sony, 18 March 2014. [Online]. Available: https://blog.playstation.com/2014/03/18/introducing-project-morpheus. [Accessed 18 8 2024].

[172] HTC, "Company Overview," [Online]. Available: https://www.htc.com/us/about/. [Accessed 4 12 2023].

[173] Nptelhrd, "Historical perspective," 2016.

[174] R. A. I. Cardoş, O. A. David and D. O. David, "Virtual reality exposure therapy in flight anxiety: A quantitative meta-analysis," *Computers in Human Behavior,* vol. 72, pp. 371-380, 2017.

[175] Statista, "Techology & Telecommunication - Hardware," Statista, [Online]. Available: https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size. [Accessed 27 9 2023].

[176] G. Makransky, T. Terkildsen and R. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and Instruction,* vol. 60, pp. 225-236, 2019.

[177] HaptX, "HaptX Gloves G1," HaptX, [Online]. Available: https://haptx.com. [Accessed 4 5 2025].

[178] J. Smed, T. Suovuo, N. Skult and P. Skult, Handbook on interactive storytelling, John Wiley & Sons, Incorporated, 2021, pp. 82-89.

[179] Amazon, "Gear VR," Amazon, [Online]. Available: https://www.amazon.com. [Accessed 4 5 2025].

[180] Zappar, "Zapbox: Mixed reality for everyone by Zappar," Zappar, [Online]. Available: https://www.zappar.com. [Accessed 4 5 2025].

[181] Skyreal, "The VR CAVE, halfway batween reality and virtuality," Skyreal, [Online]. Available: https://skyreal.tech/news/the-vr-cave-halfway-between-reality-and-virtuality. [Accessed 4 5 2025].

[182] Lumoplay, "LumoPlay Make any display interctive," Lumoplay, [Online]. Available: https://www.lumoplay.com/. [Accessed 4 5 2025].

[183] Igloo, "Unlock the potential of an Immersive Space in Education," Igloo, 19 12 2023. [Online]. Available: https://www.igloovision.com/about/blog/immersive-classroom. [Accessed 4 5 2025].

[184] L. Edutech, "Immersive Interactive Classroom," Youtube, 2023.

[185] PlayStation, "PlayStationVR," PlayStation, [Online]. Available: https://www.playstation.com/fi-fi/ps-vr/. [Accessed 4 5 2025].

[186] bigscreen, "This is Beyond 2," Bigscreen, [Online]. Available: www.bigscreenvr.com . [Accessed 4 5 2025].

[187] Shiftall, "MeganeX," Shiftall, [Online]. Available: https://en.shiftall.net/products/meganex. [Accessed 4 5 2025].

[188] Meta, "Meta Quest," Meta, [Online]. Available: https://www.meta.com/fi/en/quest/. [Accessed 4 5 2025].

[189] H. Vive, "Vive XR Elite," HTC Vive, [Online]. Available: https://www.vive.com/us/product/vive-xr-elite/overview/. [Accessed 4 5 2025].

[190] Pico, "Pico 4," Pico, [Online]. Available: https://www.picoxr.com/global/products/pico4. [Accessed 4 5 2025].

[191] D. Murray, Interaction design, London: Goldsmiths University of London, 2010.

[192] T. Carey, "The gift of good design tools," in *Advances in Human-Computer Interaction, Volume 2*, Ablex, Norwood, NJ, 1988, pp. 175-213.

[193] Y. Rogers, H. Sharp and J. Preece, Interaction design - beyond human-computer interaction, Glasgow: Bell & Bain Ltd, Wiley, 2016.

[194] L. Hallnäs and J. Redström, Interaction design, foundations, experiments, Borås: The Interactive Institute, The Swedish School of Textiles, University College of Borås, 2006.

[195] A. Dix, "Human-Computer Interaction," in *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*, 2018, p. 1327.

[196] A. Dix, "Human-Computer Interaction," in *Encyclopedia of Database Systems*, Boston, Springer, 2009, pp. 1327-1331.

[197] T. D. Cook and D. T. Campbell, Quasi-Experimentation Design & Analysis Issues for Field Settings, U.S.A: Rand McNally College, 1979, p. 18.

[198] E. Adams, Fundamentals of game design (3rd ed.), USA: Pearson Education, 2014, pp. 259-260, 283, 288-289.

[199] J. Schell, The Art of Game Design : A Book of Lenses, Second Edition, Second edition. Boca Raton : A K Peters/CRC Press, 2015, pp. 210, 211.

[200] D. Dowling, C. O. Fearghail, A. Smolic and S. Knorr, "Faoladh: A Case Study in Cinematic VR Storytelling and Production.," in *Rouse, R; Koenitz, H; Haahr, M (eds) Interactive storytelling. ICIDS 2018. Lecture notes in Computer Science, vol 11318*, Dublin, 2018.

[201] P. Ehn and R. Badham, "Participatory Design and Collective Designer," *Proceedings of the Participatory Design Conference,* pp. 23-25, 2002.

[202] J. Gregory, "Scandinavian Approach to Participatory Design," *International Journal of Engineering Education,* vol. 19, no. 1, pp. 62-74, 2003.

[203] Bjørn-Andersen and B. Hedberg, "Designing information systems in an organisational perspective," *Studies in the Management Science: Perscriptive Models of Organisations,* vol. 5, pp. 125-142, 1977.

[204] G. Bjerknes and T. Bratteteig, "User Participation and Democracy: A Discussion of Scandinavian Research on System Development," *Scandinavian journal of Information Systems,* vol. 7, no. 1, pp. 73-98, 1995.

[205] G. Bjerknes and T. Bratteteig, "User Participation and Democracy: A Discussion of Scandinavian Research on System Development," *Scandinavian Journal of Information Systems,* vol. 7, no. 1, pp. 73-98, 1995.

[206] M. Nouwen, S. Schepers, K. Mouws, K. Slegers, K. N and P. Duysburgh, "Designing an educational music game: What if children were calling the tune?," *International Journal of Child-Computer Interaction,* vol. 9, no. 10, pp. 20-32, 2016.

[207] M. Rice and L. Valdivia, "A Simple Guide for Design, Use and Evaluation of Educational Materials," *Health Education Quarterly,* vol. 18, no. 1, pp. 79-85, 1991.

[208] R. Molic and J. Nielsen, "Improving a Human-Computer Dialogue," *Communications of the ACM,* vol. 33, no. 3, p. 338, 1990.

[209] J. Nielsen, "Enhancing the Explanatory Power of Usability Heuristics," *Human Factors in Computing Systems CHI '94 "Celebrating Interdependence",* pp. 152-158, 1994.

[210] J. Nielsen, "10 Usability Heuristics for User Interface Design," Nielsen Norman Group, 30 January 2024. [Online]. Available: https://www.nngroup.com/articles/ten-usability-heuristics/. [Accessed 15 July 2024].

[211] J. Nielsen, "Severity Ratings for Usability Problems," NN/g, 1 11 1994b. [Online]. Available: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/. [Accessed 8 8 2025].

[212] R. Prieto de Lope, N. Medina-Medina, P. Paderewski and F. L. Gutiérrez-Vela, "Design methodology for educational games based on interactive screenplays," *CEUR Workshop Proceedings.,* 2015.

[213] E. Dogan and F. Sahin, "Mapping the Research Agenda in Virtual Reality Studies within Education," *Turkish Online Journal of Distance Education,* vol. 25, no. 3, p. 265–293, 2024.

[214] I. 25000b, "The ISO/IEC 25000 series of standards," ISO, [Online]. Available: https://iso25000.com/index.php/en/iso-25000-standards. [Accessed 13 7 2025].

[215] C. M. Judd, E. R. Smith and L. H. Kidder, Research methods in social relations (6th ed.), New York: Holt, Rinehart and Winston, 1991.

[216] S. Estonia, "Modernisation of data collection for social surveys and adaptation with pandemic crisis.," 27 9 2021. [Online]. Available: www.stat.ee. [Accessed 23 6 2025].

[217] R. Dulaney, "Westat," 1 5 2024. [Online]. Available: https://census.gov/fedcasic/fc2024/pdf/7aVideoInterviewingonMEPS.pdf. [Accessed 23 6 2025].

[218] N. Schwartz, B. Knäuber, D. Oyserman and C. Stich, "Psychology of asking questions," in *International handbook of survey methodology*, New York, Taylor & Francis, 2008, pp. 25-26.

[219] Webropol, "Survey and reporting," Webropol Group, [Online]. Available: https://webropol.co.uk/survey-and-reporting/. [Accessed 8 July 2024].

[220] M. T. Roberson and E. Sundstrom, " Questionnaire Design, Return Rates, and Response Favorableness in an Employee Attitude Questionnaire," *Journal of applied psychology,* vol. 75, no. 3, pp. 354-357, 1990.

[221] M. Kalliopuska, Psykologian sanasto, Keuruu: Otavan Kirjapaino Oy, 2005, p. 117.

[222] P. Bartram and D. Yielding, "The Development of an Empirical Method of Selecting Phrases Used in Verbal Rating Scales: A Report on a Resent Experiment," *Journal of the Market Research Society,* vol. 15, pp. 151-156, 1973.

[223] R. M. Worcaster and T. R. Burns, "A Statistical Examination of the Relative Precision of Verbal Scales," *Journal of the Market Research Society,* vol. 17, pp. 181-197, 1975.

[224] A. R. Wildt and M. B. Mazis, "Dominants of Scale Response: Label Versus Position," *Journal of Marketing Research,* vol. 15, pp. 261-267, 1978.

[225] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology,* vol. 22, no. 140, pp. 5-55, 1932.

[226] Meta, "Facebook," Meta, [Online]. Available: https://www.facebook.com. [Accessed 12 July 2024].

[227] J. Cohen, Statistical power analysis for the behavioral science, New York: Lawrence Erlbaum assiciates, 1988.

[228] D. George and P. Mallery, IBM SPSS Statistics 23 Step by Step, A Simple Guide and Reference, fourteenth edition, New York: Routledge, 2016.

[229] I. 25000, "ISO/IEC 25010," ISO, [Online]. Available: https://iso25000.com/index.php/en/iso-25000-standars/iso-25010. [Accessed 13 7 2025].

[230] A. Silva, P. Pinheiro, A. Albuquerque and J. Barroso, "A Process for Creating the Elicitation Guide of Non-Functional Requirements," in *Sofware Engineering Perspectives and Application in Intelligent Systems. ITCTIS CSOC 2017 2016. Advances in Intelligent Systems and Computing, vol 465, Springer, Cham.*, 2016.

[231] R. B. Bias and D. J. Mayhew, Cost-Justifying Usability, An update for the Internet age (Second Edition), Morgan Kaufmann, 2005.

[232] Sidequest, "Sidequest VR," [Online]. Available: https://sidequestvr.com/.

[233] U. engine, "Features," UNREAL ENGINE, [Online]. Available: https://www.unrealengine.com/. [Accessed 15 11 2024].

[234] Unity, "Unity Resources," Unity engine, [Online]. Available: https://unity.com. [Accessed 15 11 2024].

[235] N. Gitter and A. DeDonato, Directors, *Principles of spatial design.* [Film]. Apple, 2024.

[236] E. Adams, Resolutions to some problems in interactive storytelling, University of Teesside, 2013, pp. 37-44, 79-82.

[237] C. Crawford, Chris Crawford on interactive storytelling, New Riders, 2013, pp. 214-223.

[238] H. Arponen, T. Häggström, S. Jortikka, M. Leinonen and T. Nyberg, Elo, Elämä, Keuruu: Otavan Kirjapaino Oy, 2017.

[239] N. B. Zeman, Storytelling for interactive digital media and video games, Boca Raton: CRC Press, 2017, pp. 228-236.

[240] R. Aylett, "Narrative in Virtual Environments - Towards Emergent Narrative," Business House, University of Salford, Salford, 1999.

[241] U. Spierling, "Adding Aspects of "Implicit Creation" to the Authoring Process in Interactive Storytelling," in *Virtual Storytelling. Using Virtual Reality Technologies for Storytelling*, 2007.

[242] B. Laurel, Computers as Theatre, USA: Addison-Wesley, 1991, pp. 99-112.

[243] N. Knoller, "Agency and the Art of Interactive Digital Storytelling," University of Amsterdam, The Neatherlands, 2010.

[244] J. Bates, "Virtual Reality, Art, and Entertainment," Carnegia Mellon University, 1991.

[245] J. Bizzocchi, "Games and Narratives: An Analytical Framework," *Loading,* vol. 1, no. 1, 2007.

[246] V. Kovalenko, M. Marienko and A. Sukhikh, "The use of augmented and virtual reality tools in a general secondary education institution in the context of blended learning," *Information Technologies and Learning Tools,* vol. 86, no. 6, pp. 70-83, 2021.

[247] S. Lee, M. Kim, H. Kim and C. Park, "Visual fatigue induced by watching virtual reality device and the effect of anisometropia," *Ergonomics,* vol. 64, no. 12, pp. 1522-1531, 2021.

[248] C. Peckmann, K. Kannen, M. Pensel, S. Lux and A. Philipsen, "Virtual reality induces symptoms of depersonalization and derealization: A longitudinal randomised control trial," *Computers in Human Behavior,* vol. 131, no. 107233, pp. 1-10, 2022.

[249] M. McCauley and T. Sharkey, "Cybersickness: Perception of Self-Motion in Virtual Environment," *Presence,* vol. 1, pp. 311-318, 1992.

[250] J. Clifton and S. Palmisano, "Effects of steering locomotion and teleporting on cybersickness and presence in HMD-based virtual reality," *Virtual Reality,* vol. 24, pp. 453-468, 2020.

[251] C. M. Barnum, Usability, Testing, Essentials, Ready, Set...Test!, Morgan Kaufmann , 2021.

[252] K. Moran, "Usability (User) Testing 101," NN/g, 1 12 2019. [Online]. Available: https://www.nngroup.com/articles/usability-testing-101/. [Accessed 2025 7 27].

[253] P. Morville and L. Rosenfeld, Information Architecture for the World Wide Web: Designing Large Scale Web Sites, 3rd ed., O'REILLY, 2006.

[254] P. Morville, "User Experience Design, Semantic Studios," 21 6 2004. [Online]. Available: https://semanticstudios.com/user_experience_design. [Accessed 22 7 2025].

[255] L. Levy and J. Novak, Game Development Essentials, Game QA & Testing, NY: Delmar Cengage Learning, 2010.

[256] R. A. Virzi, "Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?," *Human Factors,* vol. 34, no. 4, pp. 457-468, 1992.

[257] J. Nielsen, "Finding usability problems through heuristic evaluation," *CHI '92,* Vols. May 3 - 7, pp. 373-380, 1992.

[258] J. Nielsen and T. K. Landauer, "A Mathematical Model of the Finding of Usability Problems," in *INTERCHI '93*, Amsterdam, 1993.

[259] T. Normaalikoulu, "Research," Univesity of Turku, [Online]. Available: https://sites.utu.fi/tnk/en/research-test-version-on-translation/. [Accessed 18 7 2025].

[260] A. Ginty, "Psychometric Properties.," in *Encyclopedia of Bahavioral Medicine*, New Yourk, Springer, 2013, pp. 1563-1564.

[261] K. Phan, T. Wager, S. Taylor and I. Liberzon, "Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI," *NeuroImage,* vol. 16, pp. 331-348, 2002.

[262] B. Salem and M. Rauterberg, "Power, Death and Love: Triology for Entertainment," in *Entertainment Computiong - ICEC 2005*, Sanda, Japan, 2005.

[263] J. VandenBerghe, "Mapping the lesson of psychology to game design," *Game Developer,* May 2012.

[264] N. D. Jackson, M. C. Ashton and J. L. Tomes, "The six-factor model of personality: facets from the big five," *Personality and Individual Differences,* vol. 21, no. 3, pp. 391-402, 1996.

[265] R. Bartle, "Virtual Worlds: Why People Play," University of Essex, 2005.

[266] Opetushallitus, "Suomi toisena kielenä ja kirjallisuus -opetus," Opetushallitus, 14 2 2025. [Online]. Available: https://www.oph.fi/fi/koulutus-ja-tutkinnot/suomi-toisena-kielena-ja-kirjallisuus-opetus-0. [Accessed 29 8 2025].

[267] C. O. Fritz and P. E. Morris, "Effect Size Estimetes: Current Use, Calculations, and Interpretation," *Journal of Experimental Psychology: General,* vol. 141, no. 1, pp. 2-18, 2012.

[268] P. Black and D. William, "Assessment and the Design of Educational Materials.," *In Educational Designer: Journal of the International Society for Design and Development in Education,* vol. 2, no. 7, 2014.

[269] L. koulu, "Perusopetuslaki muuttuu: Liikunnallisen elämäntavan edistäminen velvoittavaksi elokuusta 2026 alkaen," 29 4 2025. [Online]. Available: https://liikuvakoulu.fi/perusopetuslaki-

muuttuu-liikunnalisen-elamantavan-edistaminen-velvoittavaksi-elokuusta-2026-alkaen/. [Accessed 1 9 2025].

[270] N. Alalwan, L. Cheng, H. Al-Samarraie, R. Yousef, A. I. Alzahrani and S. M. Sarsam, "Challenges and Prospects of Virtual Reality and Augmented Reality Utilization among Primary School Teachers: A Developing Country Perspective," *Studies in Educational Evaluation,* vol. 66, 2020.

[271] Zoom, "Create focused discussion using breakout rooms," [Online]. Available: https://www.zoom.com/en/products/virtual-meetings/features/breakout-rooms/. [Accessed 22 9 2025].

[272] C. F. Hasibuan, B. Haetono and T. Wijayanto, "Effect of prior gaming experience on cybersickness symptoms in a virtual reality environment," *Virtual Reality & Intelligent Hardware,* vol. 7, no. 5, pp. 468-482, 2025.

[273] P. A. Suri, M. A. Syahputra, A. S. H. Ahmad Syauqy Hilmi Amany and A. Djafar, "Systematic literature review: The use of virtual reality as a learning media," *Procedia Computer Science,* vol. 216, pp. 245-251, 2022.

[274] E. Gene, F. N. Cakmak, H. Ciftci and Z. M. Hocaoglu, ""Fiction is the reality": A qualitative study on digital game addiction and reality perception in young adults," *Children and Youth Services Review,* vol. 157, no. 107445, pp. 1-6, 2924.

[275] U. U. Kaya, M. S. Koyuneu, M. Arabaci and N. H. Karaca, "Exploring latent profiles of digital game addiction, socio-emotional well-being and resilience in Turkish preschoolers," *Acta Psychologica,* vol. 261, pp. 1-9, 2025.

[276] I. Gorbanev, S. Agudelo-Londoño, R. González, A. Cortes, A. Pomares, V. Delgadillo, F. Yepes and Ó. Muñoz, "A systematic review of serious games in medical education: quality of evidence and pedagogical strategy," *Med Educ Online,* vol. 23, no. 1, 2018.

[277] L. Wang, B. Chen, G. Hwang and Y. Wang, "Effects of digital game-based STEM education on students’ learning achievement: a meta-analysis," *International Journal of STEM Education,* vol. 9, no. 26, 2022.

[278] Y. Zhonggen, "A Meta-Analysis of Use of Serious Games in Education over a Decade," *International Journal of Computer Games Technology,* vol. 2019, 2019.

[279] A. All, E. N. P. Castellar and J. Van Looy, "Digital Game-Based Learning effectiveness assessment: Reflections on study design," *Computers & Education,* vol. 167, 2021.

[280] D. Clark, E. Tanner-Smith and S. Killingsworth, "Digital Games, Design, and Learning:A Systematic Review and Meta-Analysis," *Review of Educational Research,* vol. 86, no. 1, pp. 79-122, 2016.

[281] M. Tienari, Tietotekniikan alkuvuodet, Helsinki: Suomen Atk-kustannus Oy, 1991.

[282] M. Ljungqvist and A. Sonesson, "Selling out Education in the Name of Digitalization: Critical Analysis of Swedish Policy," *Nordic Jornal of Studies in Educational Policy,* vol. 8, no. 2, pp. 89-102, 2022.

Appendices

Appendix 1. Questionnaire for teachers

TURUN YLIOPISTO

VR-oppimateriaalit

ⓘ Pakolliset kysymykset merkitty tähdellä (*)

Tämä kysely on osa Gradu-tutkielmaani, jonka tarkoituksena on selvittää, millaista asioista hyvä virtuaalitodellisuus (VR) -oppimateriaali koostuu. Vastausten pohjalta tehdään yläkoulun biologian opetukseen soveltuva VR-oppimissovellus, jonka käytettävyyttä testataan yhden oppilasryhmän kanssa.

Tutkimukseen voivat osallistua biologian aineenopettajat sekä biologian aineenopettajaksi opiskelevat. Kyselyyn vastaaminen ei edellytä kokemusta VR:stä. Tutkimukseen vastaaminen kestää 7–15 min. On tärkeää, että mahdollisimman moni vastaa kyselyyn ja opettajien mielipide saadaan hyvin esille.

Vastaukset käsitellään anonyymisti, eikä niitä voida yhdistää vastaajaan. Vastaus ja sen kopio säilytetään Turun yliopiston levypalvelimella (seafile.utu.fi). Mahdollisesti annetut sähköpostiosoitteet ja nimitiedot poistetaan tutkimuksen päättymisen jälkeen, eikä niitä käytetä mihinkään muuhun tarkoitukseen. Jos sinulla on tutkimukseen liittyen kysymyksiä, ota yhteyttä: mkkyyh@utu.fi

! Jokainen tutkimukseen osallistunut saa halutessaan linkin, jota kautta voi ladata itselleen tutkielman tuloksena syntyneen oppimissovelluksen käytettäväksi oppilaiden kanssa biologian tunnilla. Kysely ohjaa sinut "Kiitos vastauksesta!" -sivun jälkeen toiseen kyselyyn, johon voit halutessasi jättää sähköpostiosoitteesi myöhemmin lähetettävää latauslinkkiä varten. Jos haluat kertoa tarkemmin ajatuksistasi ja/tai kokemuksistasi VR-oppisovelluksiin liittyen voit jättää yhteystietosi mahdollista haastattelua varten samalla sivulla.

Seuraava

Taustakysymykset

1. Sukupuoli: *

- ◯ Nainen
- ◯ Mies
- ◯ Muu
- ◯ En halua kertoa

2. Ikäryhmä: *

- ◯ Alle 30-vuotiaat
- ◯ 31–40 v.
- ◯ 41–50 v.
- ◯ 51–64 v.
- ◯ 65 vuotta täyttäneet

3. Opetuskokemus: *

- ◯ Alle 1 vuosi
- ◯ 2–5
- ◯ 6–10
- ◯ 11–15
- ◯ 16–20
- ◯ yli 20 vuotta

4. Opetuskokemus yläkoulussa: *

- ◯ Alle 1 vuosi
- ◯ 2–5
- ◯ 6–10
- ◯ 11–15
- ◯ 16–20
- ◯ yli 20 vuotta

5. Käytän sähköistä opetusmateriaalia (esim. digitaalinen kirja): *

- ◯ joka päivä
- ◯ muutaman kerran viikossa
- ◯ kerran viikossa
- ◯ harvemmin
- ◯ en käytä lainkaan

Edellinen Seuraava

Taustakysymykset: virtuaalitodellisuus

6. Tiedän mitä virtuaalitodellisuus (VR) on. *

- ◯ Kyllä
- ◯ Ei
- ◯ En osaa sanoa

Edellinen Seuraava

Virtuaalitodellisuudella (VR) tarkoitetaan sähköisesti luotua kolmiulotteista kuvaa, joka heijastetaan VR-lasien näytölle tai esimerkiksi huoneen seinille. Näin luodaan vaikutelma, että käyttäjä olisi fyysisesti jossain toisessa paikassa missä hän oikeasti on. VR-maailmassa kaikki mitä nähdään on keinotekoisesti luotua ja käyttäjälle annetaan usein mahdollisuus vuorovaikuttaa ympäristössä eli esimerkiksi ottaa jokin esine käteen tai liikkua. Liikkuminen tapahtuu joko fyysisesti kävelemällä pelialueen rajojen sisäpuolella tai "teleportaamalla" eli kaukosiirtymällä (hyppimällä) virtuaalimaailman sisällä.

Virtuaalitodellisuuslasit käytössä

Edellinen Seuraava

7. Oma kokemus virtuaalitodellisuudesta: *

- ◯ olen käyttänyt useasti virtuaalilaseja
- ◯ olen käyttänyt muutaman kerran
- ◯ en ole käyttänyt virtuaalilaseja

8. Kokemus virtuaalitodellisuudesta oppilaiden kanssa: *

- ◯ olen käyttänyt useasti oppilaiden kanssa virtuaalilaseja
- ◯ olen käyttänyt muutaman kerran oppilaiden kanssa
- ◯ en ole käyttänyt virtuaalilaseja oppilaiden kanssa

9. Monetko lasit on ollut käytössä? *

[Valitse ⌄]

10. Minkä valmistajan lasit ja mikä malli? (voit valita useita vaihtoehtoja) *

- ☐ Oculus Rift S
- ☐ Oculus Quest 2
- ☐ HTC Vive Cosmos
- ☐ HTC Vive Pro
- ☐ Valve Index
- ☐ Microsoft HP Reverb
- ☐ muu , mikä: ____________________
- ☐ en osaa sanoa

11. Mitä sovelluksia olet(te) käyttäneet? *

12. Kerro tarkemmin kokemuksistasi VR-lasien käytössä.

Edellinen Seuraava

Oppisisältö ja opettajankokemus

13. Miten tärkeänä pidät seuraavia väittämiä VR-oppimateriaaleissa?

1 = täysin eri mieltä, 2 = jokseenkin eri mieltä, 3 = jokseenkin samaa mieltä, 4 = täysin samaa mieltä *

	1	2	3	4	en osaa sanoa
A. Materiaalin sisältö sopii opetussuunnitelmaan *	○	○	○	○	○
B. Materiaalin käyttö vaatii opettajalta vähän etukäteisvalmisteluja *	○	○	○	○	○
C. Materiaali ottaa erilaiset oppijat huomioon *	○	○	○	○	○
D. Oppilas saa palautetta suoriutumisestaan *	○	○	○	○	○
E. Sovelluksessa on tarjolla apua ja tukea sitä tarvitsevalle *	○	○	○	○	○
F. Oppilaat oppivat käyttämään VR-laseja *	○	○	○	○	○
G. Opettaja pystyy seuraamaan oppilaiden etenemistä ja edistymistä VR-ympäristössä *	○	○	○	○	○

H. Opettaja toimii aktiivisessa roolissa ohjaajana *	○	○	○	○	○
I. Materiaali syventää oppilaiden tietoutta opetettavasta aiheesta *	○	○	○	○	○
J. Materiaali opettaa tutkivaa oppimista *	○	○	○	○	○
K. Materiaali helpottaa oppilaiden abstraktien käsitteiden ymmärrystä *	○	○	○	○	○
L. Oppilas oppii syy- ja seuraussuhteita *	○	○	○	○	○
M. Oppilas oppii biologiassa tärkeitä tiedonhankinta menetelmiä *	○	○	○	○	○
N. Oppilas oppii soveltamaan biologista tietoa *	○	○	○	○	○
O. Sovellus tarjoaa rauhallisen oppimisympäristön *	○	○	○	○	○
P. Opettaja saa positiivisen kokemuksen *	○	○	○	○	○

14. Mitä muuta haluaisit sanoa VR-oppisisältöihin ja opettajan kokemukseen liittyen?

Edellinen | Seuraava

Oppilaan kokemus

15. Miten tärkeänä pidät seuraavia väittämiä VR-oppimateriaaleissa?

1 = täysin eri mieltä, 2 = jokseenkin eri mieltä, 3 = jokseenkin samaa mieltä, 4 = täysin samaa mieltä *

	1	2	3	4	en osaa sanoa
A. Oppilaat voivat kokea uusia asioita sivusta seuraamisen sijaan *	○	○	○	○	○
B. Oppilaat voivat seikkailla ja tutkia ympäristöä *	○	○	○	○	○
C. Oppilaat voivat käyttää luovuuttaan *	○	○	○	○	○
D. Oppilaat saavat uusia elämyksiä *	○	○	○	○	○
E. Oppilas pystyy toimimaan itsenäisesti VR-ympäristössä *	○	○	○	○	○
F. Oppilas voi edetä ympäristössä omaa tahtia *	○	○	○	○	○
G. Sovellus rohkaisee oppilasta toimimaan ja kokeilemaan asioita *	○	○	○	○	○

F. Oppilas voi edetä ympäristössä omaa tahtia *	○	○	○	○	○
G. Sovellus rohkaisee oppilasta toimimaan ja kokeilemaan asioita *	○	○	○	○	○
H. Oppilaat voivat tehdä yhteistyötä ja pohtia etenemistään ja ratkaisuja yhdessä *	○	○	○	○	○
I. Oppilaat voivat nähdä toisensa virtuaaliympäristössä *	○	○	○	○	○
J. Oppilaat voivat nähdä opettajan virtuaaliympäristössä *	○	○	○	○	○
K. Oppilaiden motivaatio oppiainetta kohtaan kasvaa *	○	○	○	○	○
L. Oppilaat voivat kokea VR-ympäristössä jotain mikä ei oikeassa ympäristössä olisi mahdollista (esim. lentää) *	○	○	○	○	○
M. Oppilaat saavat positiivisen oppimiskokemuksen *	○	○	○	○	○

16. Mitä muuta haluaisit sanoa VR-oppimateriaaleihin ja oppilaan kokemukseen liittyen?

Edellinen | Seuraava

Helppokäyttöisyys ja tekniset ratkaisut

17. Miten tärkeänä pidät seuraavia väittämiä VR-oppimateriaaleissa?

1 = täysin eri mieltä, 2 = jokseenkin eri mieltä, 3 = jokseenkin samaa mieltä, 4 = täysin samaa mieltä *

	1	2	3	4	en osaa sanoa
A. Sovellus on helppokäyttöinen *	○	○	○	○	○
B. Sovellus sisältää tutoriaalin, joka opastaa sovelluksen käyttöön ensimmäisellä kerralla *	○	○	○	○	○
C. Oppilaiden pääsy muihin sovelluksiin on rajattu. *	○	○	○	○	○
D. Sovelluksen ulkonäkö on esteettisesti miellyttävä *	○	○	○	○	○
E. Sovelluksessa on näkyvillä vain kulloinkin tarvittavat toiminnot *	○	○	○	○	○
F. Sovelluksen ympäristö on todellisen maailman kaltainen *	○	○	○	○	○

G. Sovelluksen käyttö on turvallista *	○	○	○	○	○
H. VR-ympäristössä liikkuminen ei aiheuta oppilaille huimausta tai pahoinvointia (motion sickness/cybersickness) *	○	○	○	○	○
I. Sovelluksen käyttö ei altista oppilaita peliriippuvuudelle *	○	○	○	○	○
J. Sovellus ei kerää käyttäjistään tarpeettomia tietoja *	○	○	○	○	○
K. Sovellusta ohjataan erillisillä käsiohjaimilla (kuva 1) *	○	○	○	○	○
L. Sovellusta ohjataan omien käsien liikkeillä (kuva 2) *	○	○	○	○	○
M. Opeteltavia ohjaus/liikkumiskomentoja on mahdollisimman vähän *	○	○	○	○	○

Kuva 1. VR-sovelluksen ohjaaminen käsiohjaimilla. Valintoja voidaan tehdä painamalla ohjaimen nappeja.

Kuva 2. VR-sovelluksen ohjaaminen käsien liikeillä. Oikenpuoleisessa kuvassa on nipistysliike, jolla voidaan esim. painaa nappuloita VR-maailmassa.

18. Mitä muuta haluaisit sanoa VR-oppisisältöihin sekä helppokäyttöisyyteen ja teknisiin ratkaisuihin liittyen?

Edellinen Seuraava

Oppitunnin pitäminen

19. Mitä asioita yläkoulun biologian tunnilla olisi mielekästä opettaa VR-sovelluksen kautta? *

20. Minkä olisi hyvä ajallinen pituus VR-oppimateriaalin läpikäynnille? *

- alle 10 min
- 11–20 min
- 21–30 min
- 31–40 min
- yli 40 min

21. Miten usein VR-oppimateriaalia tulisi biologian opetuksessa käyttää? *

- ◯ joka oppitunti
- ◯ joka toinen tai kolmas oppitunti
- ◯ 5-10 kertaa vuosiluokan aikana
- ◯ muutaman kerran vuosiluokan aikana
- ◯ kerran vuosiluokan aikana
- ◯ ei lainkaan, syy: ____

22. Monetko virtuaalilasit olisi hyvä olla oppitunnilla 20 oppilaan ryhmässä? *

- ◯ 1-2 lasit
- ◯ 3-5
- ◯ lasit joka toiselle
- ◯ lasit kaikille oppilaille ja opettajalle

23. Mitä muuta haluaisit sanoa VR-oppitunnin pitämiseen liittyen?

[Edellinen] [Seuraava]

24. Onko vielä jotain muuta mitä haluaisit sanoa tähän tutkimukseen liittyen?

[Edellinen] [Lähetä]

Appendix 2. Questionnaire for students

VR-oppimispelin testaus

Tämä kysely on osa Gradu-tutkielmaani, jonka tarkoituksena on selvittää, millaista asioista hyvä virtuaalitodellisuus (VR) -oppimateriaali koostuu. Tämä on tutkimukseni toinen osa. Ensimmäisessä osassa tein kyselyn biologian opettajille VR:n käytöstä opetuksessa. Tämän VR-sovelluksen, jota juuri testasitte, tekemisessä on käytetty hyväksi opettajien vastauksia. Testauksen tekee yksi luokkaryhmä.

Tämä kysely on tarkoitettu VR-sovelluspeliäni testanneille.

Vastaukset käsitellään anonyymisti, eikä niitä voida yhdistää vastaajaan. Vastaus ja sen kopio säilytetään Turun yliopiston levypalvelimella (seafile.utu.fi). Jos sinulla on tutkimukseen liittyen kysymyksiä, ota yhteyttä: mkkyyh@utu.fi

Lasien numero: ____________

Kellon aika: ______________

Taustakysymykset:

Viimeisin biologian arvosanani on:					
Oletko käyttänyt VR-laseja ennen tätä testausta?	_____en ole _____muutaman kerran _____ usein Jos olet, niin mitkä lasit:____________________________				
1=täysin eri mieltä 2=jokseenkin eri mieltä 3=jokseenkin samaa mieltä 4=täysin samaa mieltä en osaa sanoa					
	1	2	3	4	eos
Miten kiinnostunut olet opiskelemaan biologiaa?					
Miten kiinnostunut olit käyttämään VR-laseja, ennen testausta?					
Miten kiinnostunut olet käyttämään niitä nyt testauksen jälkeen?					

Kokemus:	1	2	3	4	eos
Piditkö VR-pelistä?					
VR-lasien käyttö toi mukavaa vaihtelua opetukseen.					
Piditkö peliin luodusta ympäristöstä.					
Oliko sinulla turvallinen olo käyttää sovellusta.					
Uskalsin kokeilla uusia asioita.					
Minulla oli kivaa.					
Tulitko huonovointiseksi?					

Hyödyllisyys:	1	2	3	4	eos
VR-peli auttoi minua oppimaan uusia lajeja.					
Saitko tarpeeksi palautetta peliltä?					
Mitä lajeja muistat pelistä:					

Mitä muita asioita muistat pelistä:

Käytettävyys ja helppokäyttöisyys:

Minun oli helppo alkaa käyttämään VR-laseja.					
Sain tarpeeksi ohjeita peliltä.					
Minua auttoi muiden ryhmäläisten ohjeet.					
Jäin jumiin enkä tiennyt mitä tehdä.					
Käsiohjainten käyttö oli helppoa.					
Saitko tehtyä mitä halusit pelissä.					
Mikä onnistui hyvin:					
Mikä oli hankalaa:					

Käyttö koulussa ja kehitysehdotukset

Miten usein haluaisit käyttää VR-laseja koulussa? _____joka oppitunti _____2-3 kertaa viikossa _____5-10 kertaa kouluvuoden aikana _____muutaman kerran kouluvuoden aikana _____kerran kouluvuoden aikana _____en lainkaan Syy:
Miten testaamaasi sovellusta voisi parantaa?
Millainen olisi täydellinen VR-oppimispeli?
Mikä virtuaalitodellisuudessa on parasta?
Mikä virtuaalitodellisuudessa on epämieluisinta?
Millaisia asioita haluaisit opiskella koulussa VR-lasien kanssa?

Mitä muuta haluaisit vielä sanoa, VR-oppimateriaaleihin ja testaukseen liittyen?