University of Tartu
Faculty of Social Sciences
Institute of Education
Curriculum: Educational Technology

Timothy Gallagher

COMPARING COLLABORATIVE PROBLEM SOLVING IN VIRTUAL REALITY AND DESKTOP COMPUTER DISPLAY ENVIRONMENTS

MA thesis

Supervisor: Senior Research Fellow of Educational Technology, Leo Aleksander Siiman, PhD

Tartu, 2018

Abstract

Comparing collaborative problem solving in virtual reality and desktop computer display environments

Although virtual reality (VR) has been used by educators for decades, due to cost, it has only recently become a relevant learning environment for educational technologists to consider using outside of military and university settings. The primary objective of this study was to compare students' use of VR as a learning environment to practice collaborative problem solving (CPS) skills with desktop activities in a primary school (n=26). The Programme for International Student Assessment (PISA) has begun measuring CPS competencies in their international assessment program, which highlights the importance of such skills. In this study, CPS related skills and attitudes towards the learning environment that were measured included self-report surveys measuring student engagement and presence, the number of words spoken during each task, the time taken to complete each task, the completion success of each task and the demonstrated actions of CPS skills in each task. Students in the VR group reported statistically significantly higher levels of engagement (M=6.15) than students in the desktop group (M=5.80). There was no statistical difference on the presence survey. There was also a significantly higher number of CPS skills demonstrated in the VR group (M=491.17) compared with the Desktop group (M=314.17). No other significant differences were found between the VR and desktop groups. This study demonstrated that certain differences between the two groups exist and that further research is needed to evaluate the benefits of using VR as a learning environment for the teaching and learning of CPS skills.

Keywords: virtual reality, collaborative problem solving, student engagement, virtual reality learning environments.

Table of Contents

Introduction

The use of virtual reality (VR) as a learning environment has been of interest to researchers for many years. This study explores differences between students completing collaborative problem solving activities in VR with desktop computers. With improvements in VR technology and the potential gains in learning outcomes that may be possible by leveraging the technology, this research sheds light on what differences may exist when students learn in VR. There is a growing volume of research that suggests VR is a valuable learning environment for various learning situations - from increasing student engagement to improving memory recall of factual information and supporting students with learning disabilities. However, little research has been done on the impact VR has on collaborative problem solving skills, which is the focus of this study.

Virtual Reality Learning Environments

Virtual Reality is a technology that can immerse a user in computer-simulated environments in which they can interact. It has the power to impact learning in extraordinary ways. Militaries have been using VR simulations for decades to train personnel in complex scenarios (Lele, 2013). Complex processes such as ocean acidification have been simulated within VR environments to help people better understand the phenomena. A range of devices exists on a broad spectrum of functionality, from low-cost options such as Google Cardboard, which leverages the technology inside an average smartphone, to high-end business and consumer devices such as the HTC Vive and Oculus Rift, which enable full-body tracking and 6 degrees of freedom headset and peripherals.

Virtual Reality has been around since at least the 1960's and is a term that has been used to describe a broad range of technologies in educational research, both in the form of hardware and software (Blanchard et al., 1990, Jensen & Konradsen, 2017, Martirosov & Kopecek, 2017,). Minecraft, Second Life and World of Warcraft are all examples of virtual worlds that have been described as VR (Jensen & Konradsen, 2017, Gregory et al., 2013). Head Mounted Displays (HMD) such as Google Cardboard and the more immersive HTC Vive are examples of VR hardware used in education (Cochrane et al., 2017, Martirosov & Kopecek, 2017). Augmented Reality (AR) and Mixed Reality (MR) are terms that describe immersive technologies that sit on the same reality spectrum (Son & Cho, 2017). Martirosov and Kopecek (2017) make a distinction between VR and AR, describing VR as a technology which simulates an environment the user occupies, while AR overlays computer-generated objects and graphics in the real environment. To put it simply, VR is more immersive than AR. For this paper, VR is defined as a technology that uses an HMD to fully immerse the user in a digitally simulated 3D environment, occluding the view of the real world (Kim et al. 2017, Martirosov & Kopecek, 2017).

Educational researchers have explored the affordances of learning in VR. Jensen and Konradsen (2017) identified VR as a useful learning environment for students to acquire cognitive skills related to remembering and understanding facts, understanding spatial and visual information and affective skills relating to controlling one's emotions and responses to stressful situations. However, they highlight significant barriers existing with current VR technology related to cybersickness, limited content, and fundamental design choices, which limit practical use in classrooms.

VR as a learning technology enables educators to develop a multitude of experiences, and it is this flexibility that is one of its significant benefits. VR simulations can be useful learning environments for education and training across a range of domains, including but not limited to, virtual field trips, emergency services, medical practice, education of students with special educational needs and collaboration and cooperation (Garcia-Cardona, Tian, & Prakoonwit, 2017, Martirosov & Kopecek, 2017). Wu et al. (2017) explored collaboration in engineering tasks and found that communication, problem-solving and spatial cognition tasks were more efficient when conducted within a VR learning environment than in a non-VR, desktop based environment. VR simulations have been used to provide medical students with opportunities to learn about internal organs which would typically require a cadaver and emergency personnel, such as firefighters, to train in simulations which would typically be too dangerous or impossible to recreate in the real world (Martirosov and Kopecek 2017). Google Expeditions, is a low-cost VR application that can simulate museums, art galleries and natural wonders such as the Great Barrier Reef. It allows teachers to take their students on virtual field trips that can provide learning opportunities for social interaction, communication and the sharing of ideas and experiences (Parmaxi, Stylianou, & Zaphiris, 2017).

Research on the application of VR with students with special educational needs has shown positive results (Lorenzo, Lledó, Pomares, & Roig, 2016). For example, Martirosov and Kopecek (2017) describe students with Autism Spectrum Disorder (ASD) benefiting from VR simulations in which they can practice skills related to social interaction. Developing simulations for people with Intellectual Disabilities, which enable them to repeatedly practice daily activities, such as visiting a supermarket, in a safe and controlled environment has proven beneficial (de Oliveira Malaquias & Malaquias, 2016).

VR has been shown to be useful for teaching coding skills. Vincur et al. (2017) observed students performing block-based programming activities using VR and traditional 2D block-programming environments with the results indicating the VR environment to be more compelling. It was claimed that the immersive nature of the VR experience was responsible for the preferred VR environment.

Simulating the classroom environment in VR has provided researchers with valuable data regarding teacher-student behaviour, which has the potential to impact teaching practice and virtual classroom design. Bailenson et al. (2008) set out to demonstrate the usefulness of VR for studying learning sciences by conducting a series of four experiments. These four experiments focused on teacher gaze behaviour, student classroom positioning by location, student classroom positioning by distance and student focus and attention.

Bailenson et al. (2008) used a VR teaching simulation to research teacher gaze behaviour, and the impact augmented information of gaze focus had on a teacher's distribution of gaze attention to virtual students. Teachers provided with augmented information describing gaze attention were more likely to distribute their gaze among virtual students than those without the augmented information.

Continuing with the research into teacher gaze and its impact on student learning, Bailenson et al. (2008), demonstrated that students learn better when they are positioned directly in front of a teacher within a VR learning environment when compared to sitting on the periphery of the classroom at an increased visual angle. Similarly, Bailenson et al. showed that students achieved greater learning outcomes when positioned close to a teacher compared to farther away.

Finally, Bailenson et al. (2008) explored the social influence of avatars and computer agents in VR learning environments and discovered that learners were able to learn better when positioned in a VR learning environment absent of any social or environmental influences. Examples of these influences used in the experiment included peers in the virtual classroom or cars driving past the virtual window in the student's peripheral view. Bailenson et al. describe the optimal classroom for learning in a didactic fashion as being one where the student learns alone in the environment. Bailenson et al. also point out that there are situations where co-learners are necessary, such as in collaborative problem-solving activities. VR learning scenarios remove barriers of the physical classroom and lesson design, such as being able to place all students in the optimal seat for learning or removing any external distractions.

Jensen and Konradsen (2017) address the weaknesses in VR learning research design by highlighting the lack of empirically valid evaluation instruments used. Rigorous evaluation instruments are fundamental when measuring learning outcomes (Jensen & Konradsen, 2017). Of the 21 articles, Jensen and Konradsen reviewed, not one conducted a randomised control study. Jensen and Konradsen (2017) also suggest that further research should focus on specific use cases of VR as opposed to whether or not it should be used in general.

Theory of VR affordances in learning

Situational Cognition Theory (SCT) is an educational theory that helps to explain the effectiveness of VR learning environments. Harman et al. (2017), in their study of SCT, examined the differences between study participants who were presented with a simulated environment using either a VR HMD or a computer desktop monitor. SCT describes the phenomena which result in a user being able to recall information related to a specific context they are situated within (Harman et al. 2017). They found that participants within the VR simulated environment were able to recall more information than those presented with the desktop monitor simulated environment and that the level of presence reported by participants in the VR simulated environment was higher than the participants presented with the desktop monitor simulated environment (Harman et al. 2017). Harman et al. (2017) go on to argue that the increased sense of immersion, through elevated levels of presence contributed to the improved recall performance.

Presence and immersion are two concepts that have guided researchers of educational technology and learning in VR. Martirosov and Kopecek (2017) define presence as “a state of consciousness, the psychological sense of being in the virtual environment” (p. 3) and argue that increased levels of presence lead to higher states of learner motivation and engagement, which have the potential to lead to greater learning outcomes. There are various ways by which VR environments can activate user presence. More realistic virtual environments lead to a user’s greater sense of presence, which is achievable by incorporating functions such as haptic feedback into the design of experiences (Martirosov & Kopecek, 2017).

The psychological phenomenon of immersion is a related concept to presence, whereby immersion is defined as a user's subjective impression of an experience being realistic for a consequential period of time (Martirosov and Kopecek, 2017). Higher levels of immersion lead to higher levels of presence (Moreno, 2002). Martirosov and Kopecek note that immersion is a somewhat subconscious experience and that immersive experiences are triggered by user engagement, both physical and emotional, with the virtual world they inhabit. Martirosov and Kopecek (2017) state that the creation of immersive VR environments provide learning opportunities through simulated activities and the common use of flight simulators are an example of a VR immersive environment where the user takes part in a simulated activity with the aim of improving their performance in the real world.

Virtual Reality and Cybersickness

Of all the potential VR has to impact learning in positives ways, challenges remain that need to be addressed. Cybersickness (CS) is one of these challenges, which is similar to motion sickness. However, it is triggered in virtual environments as opposed to real ones. Users suffering from CS experience symptoms such as nausea, headaches and eyestrain (Martirosov & Kopecek, 2017, Melo, et al., 2017). Iskenderova et al. (2017) describe four factors which can contribute to CS: Individual factors; Device factors; Task factors; Miscellaneous factors. Results from a study conducted by Park et al. (2006) suggest that females are more susceptible to CS within non-VR driving simulations, which is an example of an Individual factor influence. Results from a study by Melo, et al. show no significant difference between genders suffering from CS in VR simulations. It should be noted that this study was limited in design, for example, participants were not exposed to VR environments for longer than seven minutes. Other studies have shown that females do suffer from more CS than men, as do adults compared with individuals under the age of 15 (Iskenderova et al. 2017). Further research on the relationship between gender and CS is needed, specifically within VR simulations. In their discussion of influencing Device factors, Iskenderova et al. highlight the importance of device computing power, with more powerful HMDs being less likely to induce CS on users, which suggests that as VR technology continues to advance, we may potentially see a reduction of users experiencing CS.

The Conceptualisation of Student Engagement

Educational researchers have long been interested in student engagement and its connection to learning outcomes. There is a significant body of literature highlighting this important link (Mitchell & Carbone, 2011) and many researchers regard engagement as a key indicator for student academic success, which is also backed up by substantial academic literature (Zyngier, 2008). This research has resulted in schools and policymakers addressing low levels of achievement through the lens of student engagement (Fredricks & McColskey, 2012). However, Zyngier (2008) claims that there is also significant contrary evidence coming out of The Organisation for Economic Co-operation and Development (OECD) research that may point to the need to develop a deeper understanding of the link between engagement and student learning outcomes (Zyngier, 2008).

There exist a number of different conceptualisations and definitions of student engagement. A common characteristic for these is that it is a multifaceted phenomenon consisting of three meta-constructs, behavioural engagement, emotional or affective engagement and cognitive engagement (Ding, Er, & Orey, 2018, Fredricks & McColskey, 2012, Mitchell & Carbone, 2011, O'Brien & Toms, 2010, Zyngier, 2008). Fredricks & McColskey argue that further research is required to determine if using a three dimensional construct most accurately describes engagement, as some researchers have added a fourth dimension. It should also be noted that although the common three dimensional model of engagement consists of behavioural, emotional and cognitive components, even within these meta-constructs of engagement, we do find different scholarly definitions (Fredricks & McColskey, 2012).

Cognitive engagement relates to a student's cognitive participation in a learning task and to the degree in which they expend mental energy or effort, such as actively reflecting upon one's learning, applying metacognitive strategies during learning and their enthusiasm for putting in effort to solve problems (Ding, Er, & Orey, 2018, Fredricks & McColskey, 2012, Mitchell & Carbone, 2011).

Behavioural engagement which often contains a social aspect of learning describes student participation and responses to learning tasks, such as answering questions or participating in group discussions (Ding, Er, & Orey, 2018, Fredricks & McColskey, 2012, Mitchell & Carbone, 2011). Alternatively, it describes how students participate more broadly in academic and social activities and the display of positive or negative conduct, in the form of supportive or disruptive behaviour (Fredricks & McColskey, 2012).

Emotional engagement refers to a students level of affective relationship or investment in learning and how they relate to their classmates and teachers or whether or not they feel like they belong to the school community (Fredricks & McColskey, 2012, Mitchell & Carbone, 2011).

Researchers from a range of disciplines have devised different methods for measuring engagement, with the most common type being self-reports, often in the form of questionnaires (O'Brien & Toms, 2010). Another method used to measure engagement is the use of performance on specified tasks, which is not strictly a measure of engagement but is arguably correlated to engagement (Carini, Kuh, & Klein, 2006). Additionally, and often in the case of measuring human engagement with technology, collecting physiological data such as heart rate, eye-gaze and mouse-clicks on a screen has been used as a measurement tool (O'Brien & Toms, 2010). Physiological data collection techniques can be intrusive; however, with recent advancements in technology, things are improving.

O'Brien & Toms (2010), developed a survey instrument designed to measure a multidimensional state of engagement, not divided into the three commonly reported areas of engagement, cognitive, and affective, but instead, a scale comprising of six factors: Focused Attention, Perceived Usability, Endurability, Novelty, Aesthetic, and Felt Environment. The result of the development of this survey instrument was a tool that can be used as a valid and reliable measure of engagement in online shopping tasks.

From the perspective of educational research, Fredricks and McColskey (2012) highlight five core measures of student engagement in their literature review of the topic: Self-reports, Experience sampling, Teacher rating of students, Interviews, Observations.

The most common of the measurement tools is the Student self-report (Fredricks & McColskey, 2012). This method of data collections asks students to respond to a set of questions which they think best describes their feelings on a topic related to engagement. These self-reports tend to be general and not subject specific and are helpful for assessing emotional and cognitive aspects of a student's level of engagement. Fredricks and McColskey claim that these are most commonly used in educational research due to their ease of use in the classroom setting, the relatively low-cost of administration and the ease of interschool data comparison. As with any subjective data set, the reliability is questionable with a risk that students answer questions dishonestly.

Experience sampling has participants responding to a self-report questionnaire at predetermined moments throughout a set period of time as prompted by an electronic device (Fredricks & McColskey, 2012). One of the reported benefits of this type of measurement tool is that it is able to elicit responses immediately. However, this method, just like the student self-report also collects subjective data and relies on participant cooperation.

A technique that removes the student from reporting their own levels of engagement is the use of teacher rating scales or checklists, which are completed by teachers based on the behaviours of individual students within a classroom (Fredricks & McColskey, 2012). This method of engagement measurement is particularly useful for younger students who are unable to access Self-reported measures. One weakness of this type of measurement is that it is not always obvious when a student is on task or off task just by looking at them, which is why it is common practice to use the Student Self-report method alongside the Teacher Ratings of Students method to explore correlations in the two measures.

Fredricks and McColskey (2012) identified structured and semi-structured interviews as another example of how school engagement has been measured in the literature. These interviews can potentially provide in depth insight into a students schooling experience. However, similar to other reporting measures discussed, the reliability of the data must be carefully considered as there may be factors that influence student responses based on the relationship between the student and the interviewer.

Measurement of student engagement using an observation protocol is also possible (Fredricks & McColskey, 2012). This protocol can be constructed so that at predetermined points in time during a lesson the observer makes notes of particular behaviours which are or are not present. Alternatively, the observation can be more narrative like, describing in detail all the behaviours that are witnessed during a lesson. The data gained from these observations have the potential to be very beneficial as they can provide a detailed view of the student’s behaviours. However, they suffer from the disadvantage of being very time consuming and may not provide the whole picture of what the student’s actual engagement levels are across all meta-constructs.

Collaborative Problem Solving as a 21st Century Skill

The Partnership for 21st Century Learning (P21) and Assessing and Teaching of 21st Century Skills (ATC21S) are two expert groups who have developed frameworks categorising 21st century skills (van Laar et al. 2017). In their work, P21 identified three types of skills: Life and Career Skills; Learning and Innovation Skills; Information, Media and Technology Skills ("P21 Framework Definitions," 2015). While ATC21S have chosen to use four broad categories of skills, to umbrella a further ten skills within each category: Ways of Thinking; Ways of Working; Tools for Working; Ways of living in the World (Assessment & Teaching of 21st Century Skills, 2012). Although these two influential groups have come to define 21st century skills differently, as have other groups doing similar work, it is possible to find a common understanding of 21st century skills. Van Laar et al. (2017) came up with the following list in their paper, which explored the literature pertaining to the topic: Collaboration, communication, digital literacy, citizenship, problem solving, critical thinking, creativity and productivity. 21st century skills differ from that of non-21st century skills in that they are related more closely to the newly developed social and economic paradigm (van Laar et al., 2017). Schools need to equip students with skills for the 21st century because the world is changing at an increasingly rapid rate and the knowledge and skills of the 20th century will not be enough for students to succeed in the future (Marzano, 2012).

Collaborative Problem Solving (CPS) has recently been added to the OECD Programme for International Student Assessment (PISA) (OECD, 2017).

The OECD (2017) defines CPS as:

> The capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution.

Cukurova et al. (2018) argues that CPS is more than just an individual solving a problem with a group of others, it represents a more complex set of skills that requires "supporting, directing, facilitating, and coordinating the thinking of others with one's own, to achieve a mutually agreed goal" (p. 2). The AT21CS framework breaks down CPS skills into two distinct skillsets, the social skillset and the cognitive skillset. Nested within these skillsets are skill indicators and elements of these skills (Siddiq & Scherer, 2017).

CPS research has explored various forms of student collaboration, such as student-student collaboration, student-teacher collaboration and student-content collaboration. In all three cases of student collaboration, there were positive and significant impacts on student learning. However, student-student collaboration was the most impactful (Siddiq & Scherer, 2017).

CPS skills are one of the 21st centuries core competencies and inseparable from the knowledge economy (Chang et al., 2017, von Davier, et al., 2017) and it is the knowledge-rich society that has made CPS skills vital for succeeding in the world of work, education and in the broader community (Siddiq & Scherer, 2017).

Siddiq and Scherer (2017) note that we are in the early stages of CPS skills assessment research, highlighting the first major effort undertaken by the Programme for International Student Assessment (PISA) to assess, internationally, CPS skills through the inclusion of CPS skills in their 2015 study. According to von Davier, et al. (2017), the development of methods for assessing CPS skills rely on a multidisciplinary approach, including, among other areas, learning science, data science and software engineering. The CPS activity outcomes being used to assess these skills should emerge from the interaction and exchange of cognitive and collaborative skills (von Davier, et al., 2017).

In their research, Chang et al., (2017) found that CPS simulations have the potential to provide students with opportunities to discuss solutions within a problem space and can, therefore, be regarded as assessment opportunities. CPS simulations provide students with opportunities to negotiate their understanding of problem tasks with group members and develop a shared understanding of how best to come to problem solutions (Chang et al., 2017). With this research, Chang et al. noted the difference in analytical reasoning skills between successful and unsuccessful groups in CPS simulations, highlighting the fact that it may be helpful for teachers to scaffold the problem solving process by providing metacognitive strategies that can help with aspects of CPS, such as working systematically and reflecting appropriately.

In the case of the CPS PISA, students were assessed on their interactions with virtual agents. The advantage of using virtual agents is that each participant receives identical stimulus and the marking is objective due to its automation. The disadvantages of using artificial collaborative partners with scripted responses stem from the fact the task is in its nature inauthentic, which may result in reduced student engagement (Siddiq & Scherer, 2017).

There are a number of challenges that face researchers and educators looking for effective assessment of CPS skills. For example, there is currently often a reliance on each individual's score being highly dependent on their partners' cognitive and collaborative ability (von Davier, et al., 2017). Additionally, the assessment methods developed for evaluating CPS skills are often expensive and time consuming to develop and, as the literature on CPS skills assessment progresses, new tasks will need to be adapted to the shared understanding of the skills to ensure best practice (von Davier, et al., 2017). Siddiq and Scherer (2017) claim that there is a need for greater research into CPS skills assessment which involves synchronous student-student collaboration with technology. The PISA Study has gone some way to address these concerns.

PISA has been assessing students individual problem solving skills since 2003. However, it was not until the 2015 study did they first attempt to assess collaborative problem solving. It has been the advancements in digital assessment technology that has enabled the assessment of these skills at such a broad scale, with the participation of 52 countries and economies (OECD, 2017). These advancements include computer based assessment along with virtual agents and automatic scoring.

The matrix of collaborative problem-solving skills for PISA 2105 breaks collaborative problem solving into three major competencies, four major individual problem solving processes and specific skills, which each are associated with individual actions, processes and strategies ("PISA 2015 Collaborative Problem-Solving Framework," 2017). The three major competencies are: Establishing and maintaining shared understanding; Taking appropriate action to solve the problem; Establishing and maintaining group organisation. It is possible to see how these major competencies interact with the four major processes and specific skills within the Matrix of Collaborative Problem Solving Skills for PISA 2015.

Table 1. Matrix of Collaborative Problem Solving Skills for PISA 2015

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(A) Exploring and understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve the problem
(B) Representing and formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describe roles and team organisation (communication protocol/rules of engagement)
(C) Planning and executing	(C1) Communicating with team members about the actions to be/being performed	(C2) Enacting plans	(C3) Following rules of engagement, (e.g. prompting other team members to perform their tasks)
(D) Monitoring and reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organisation and roles

(Table 1 “PISA 2015 Collaborative Problem-Solving Framework,” 2017)

Consensus building, Jigsaw problems and Negotiations are three distinct CPS task-types which describe how problems are designed with specific outcomes in mind, aimed to invoke particular problem-solving interactions and behaviours (“PISA 2015 Collaborative Problem-Solving Framework,” 2017). Consensus building provides opportunities for groups to agree upon certain decisions which are not immediately obvious and require the sharing of perspectives, discussion and potential concession to be made to reach agreement. Jigsaw problems require the pooling of individual information between group members in order for solutions to be found, which ensures interdependence between the individuals. Negotiation task-types create challenging group dynamics with each individual having different personal goals, yet collective group goals and only through negotiation and mutual benefit decision making can solutions be found that satisfy all group members. It should be noted that these three CPS task-types are not an exhaustive list and that other task types can be appropriate, as long as they have similar characteristics. Additionally, each of these a single assessed activity may contain more than one task-type.

Aims of the study and research questions

The potential for VR to transform the way we learn has long been discussed. Widespread adoption has been held back by high cost and accessibility to a broad market and limited technical capabilities. Since the development of the Oculus Rift Development Kit 1, which was crowdfunded via Kickstarter, VR has entered a new phase of development and use. Although the beginnings of this phase had a greater impact on gaming and commercial use, VR learning environments are starting to be seen as serious options for educators and academics. The decades worth of academic literature exploring VR learning environments have now begun to be updated by contemporary studies along with the latest hardware. Although a large number of these studies lack generalisability due to methodology and sample size, the new phase of VR research is maturing at a healthy rate.

What the current literature does tell us is that VR has the ability to immerse its users in artificial environments. This feeling of presence, triggered by a well designed virtual environment, can lead to immersion and increased levels of engagement. Through increased levels of student engagement, research shows that we are also able to increase levels of achievement and student outcomes.

This new round of VR research has only explored in a limited capacity the capabilities VR has in relation to Collaborative Problem Solving Skills. Computer desktop systems are used as the learning environment for the administration of PISA CPS activities for students. While this has been successful in their first iteration of the international examination, the technology being used is somewhat dated.

The purpose of this research is to explore ways in which educators can use VR learning environments as a means to teach CPS skills. The research question is: To what degree do students differ in completing CPS activities when using VR technology compared to completing CPS activities using a desktop computer?

Methods

Participants

This study was designed to investigate the differences between collaborative problem solving activities in virtual reality compared with desktop activities. The study population (n = 105) were in their final year of primary education at a bilingual, Chinese international school located in Shanghai, China. The ages of the population ranged between 11 and 12 years *(M=11.25, SD = 0.433)*. As the aim of this research was to study the average student from within the population, the sampling strategy employed selected students who achieved an average score on an externally assessed, adaptive standardised test, which assessed students Developed Ability. Developed Ability is a combination of Verbal and Non-Verbal reasoning skills. A total of 56 students fell within the average distribution range and of these 56 students, 24 students were randomly selected to take part in the experiment. There was an equal number of males to females assigned to the final sample. These 24 students were then randomly assigned to either the Virtual Reality (VR) group or the Desktop (DT) group. Each of these groups contained 12 students. The VR group consisted of 5 female and 7 male students, while the DT group consisted of 7 female and 5 male students. Students assigned to either the VR Group or DT Group were then randomly assigned a partner with whom they would complete the CPS tasks. Of the VR group pairs, there were three female/male pairs, two female/female pairs and one male/male pair. Of the DT Group, there were three female/male pairs, two male/male pairs and one female/female pair.

Research Design

The VR and DT groups were assigned the same three collaborative problem-solving tasks. However, each group was required to interact with the tasks using different technology. Before beginning the problem-solving tasks, each pair was introduced to the activity with a short PowerPoint presentation. The introduction was conducted in the form of a mini-lesson and outlined the learning intention for the lesson, “Learning Intention: To work with my partner to solve problems”, a brief overview of the three problem solving task scenarios, and a matrix of the CPS skills that they would be practising. The introduction to problem solving occurred outside of VR for both groups.

The VR Group had an additional component to the experiment sequence which provided the students with a chance to learn the navigation and interaction controls of the software. This component was not necessary for the DT Group as the software used by the students was very familiar.

The first problem solving task type was a Consensus building task, which required each pair of students to reach consensus on a set of awards for eight different pieces of Harry Potter themed artwork. These awards were the prizes of 1st through 7th place. The students were asked to consider 3 criteria for awarding the prizes: Creativity, Imagination and Skill. The second and third problem solving task types were both Jigsaw problems, which required students to determine a four-digit pin code by sharing information of which only they had access. In the first instance, one member of the pair had images of four playing cards, while the other member of the pair had instructions for how to order these playing cards. In the second instance, the roles were reversed with the student who initially had the playing cards now had the set of instruction cards. In both the VR and DT group instances, the students were only able to communicate verbally and were unable to see each other while collaboratively solving the problem.

The order by which the tasks were presented to the students was the same for both the VR group and DT group. The Consensus building task was presented first, followed by the two Jigsaw problem tasks. In the case of the VR group, each task was contained within a single environment, with the instructions for how to complete the tasks along with 1st through 7th place award cards and artwork. The task was spread out over 8 PowerPoint slides for the DT group. The first and second slide included the name of the activity while the third through sixth slide displayed two pieces of artwork respectively. The instructions and artwork were identical for each group.

Technology

Each member of the VR Group interacted with the virtual environment through an HTC Vive, which is a high-end consumer VR device, offering room-scale positional tracking of the hands and head. Participants were able to teleport within the VR scene using the hand controllers as well as manipulate the resource cards and paintings by size and position. The DT Group interacted with the environment on a laptop screen and were provided with a pen and paper to record their rankings of the artwork as well as the pin code solution.

Measurement

Six areas were measured to answer the research question: To what degree do students differ when completing CPS activities when using VR technology compared a desktop computer?

1. Immersion
2. Engagement
3. Number of words spoken during activity
4. Time to complete task
5. Completion Success
6. Demonstrated actions of Collaborative Problem Solving skills.

Immersion - Immersive Self-Report Survey

Parong & Mayer's (2018), 7-point Likert-type scale immersive self-report survey was adapted to record students interest, motivation, engagement, and affective states for comparison between the two study groups. This survey was initially designed for comparing two virtual reality learning interventions. As the initial use case for this survey was to compare two virtual reality learning interventions, changes were made to make it suitable for comparing a VR group and DT group. Appendix 1 displays the Paraong & Mayer's (2018) survey, the changes made for this experiment and the reasons for changes.

Engagement - Engagement Self-Report Survey

To measure and compare levels of engagement between the VR and DT group, an adapted version of O'Brien and Toms (2010) engagement self-report survey was used. This survey was initially created to collect user engagement in online shopping environments. It was tested for reliability and validity and aimed to measure six engagement factors: Perceived Usability, Aesthetics, Novelty, Felt Involvement, Focused Attention, and Endurability. The changes to the survey were made to ensure it was applicable to the nature of the experiment. Not all questions from the original survey were included, and questions were reworded to match the context of the experiment. The changes to the survey and the rationale for these changes can be found in Appendix 2.

Time to complete task

There were three different measurements used to compare the VR and DT groups for time to complete the task. The amount of time to complete the Consensus activity, the amount of time to complete both Jigsaw activities and overall time, which included both the Consensus and Jigsaw activities.

Number of words spoken

The number of words spoken was recorded by transcribing the spoken interactions between students. Any non-word, such as, um or ah, were not included in the total word count. Just like the Time to complete task measurement, the number of words spoken measurement was collected in three batches, the numbers of words spoken during the Consensus activity, the number of words spoken during the two Jigsaw activities and the overall number of words.

There were some challenges with recording the number of words spoken due to the quality of audio recording. On occasions when students did not pronounce words clearly enough for the audio recording to pick up the meaning, the word was recorded as [inaudible] in the transcript. All [inaudible] words were recorded as one word spoken.

Completion Success

The success of the groups to complete the Jigsaw activities were measured. The Jigsaw activities had demonstrably correct answers, while the Consensus activity had no single correct answer and was therefore not included in the Completion Success measure.

Demonstrated Actions of Collaborative Problem Solving Skills

The assessment rubric from the PISA Framework for Collaborative Problem Solving was used to record the CPS actions of each student. This rubric breaks each CPS action into two dimensions, collaborative competence and the problem solving process. Each group's session was recorded by video and then re-watched and assessed using the rubric.

Pilot Groups

Before the commencement of the primary study, one pilot group was used to test the suitability of the measurements. There were lessons learned from this pilot group which influenced the final methods of measurement. Firstly, the Engagement and Immersion surveys were rewritten to ensure the language was easily understandable for the students reading ability. There were still occasions where participants asked the researcher about the meaning of individual words in the questions. In these cases, the researcher provided additional context to clarify the meaning. Secondly, the VR pilot group was initially able to move within the physical environment on a 2m x 1.5m space which translated to the virtual environment. For safety reasons, students in the research group completed the activity seated in the centre of the virtual environment. Finally, the researcher introduced the experiment sequence, which was a short task providing the participants with the chance to learn and experiment with the VR controls. There was no need to adapt the DT Group experiment, and so the data collected during the DT Group pilot was included in the study.

Inter-rater Reliability

A sample of 64 discourse messages were independently coded by two persons according to the aforementioned CPS rubric. The inter-rater reliability of the coding was calculated using Cohen's kappa coefficient. The Cohen's kappa coefficient for the coding of the Collaborative Competence (CC) strand of the rubric was 0.545. The Cohen's kappa coefficient for the coding of the Problem Solving Process (PSP) was 0.449. The Cohen's kappa for the CPS skills, which are CC and PSP combined, was 0.503. The results of the Cohen's kappa coefficient indicate that there was weak inter-rater reliability (McHugh, 2012). The limitations of the CPS rubric and the conclusions that can be drawn as a result of these data have been addressed in the discussion section.

Results

This study aimed to answer one central research question: To what degree do students differ in completing CPS activities when using VR technology compared to completing CPS activities using a desktop computer?

The following section presents the results of the two questionnaires examining participant presence and engagement, the number of words each group spoke during the CPS activities, the amount of time each group took to complete the CPS activities, the successful completion rate of the two Jigsaw tasks and the number of demonstrated Collaborative Problem Solving instances. The results are evaluated using the Mann-Whitney U Test. The purpose of this study was to compare the results of the Virtual Reality group with the Desktop group completing the same collaborative problem solving tasks.

Table 1. Results of Engagement Survey given to students after completing the CPS tasks.

Engagement survey item	VR group *M (SD)*	DT group *M (SD)*	U-statistic	p-value
1. I used a lot of mental effort in the lesson	5.17 (1.67)	5.50 (1.04)	68.5	0.865
2. I felt that the activity was difficult	2.58 (1.55)	2.33 (1.24)	67.5	0.818
3. I enjoyed learning this way	6.83 (0.37)	6.33 (0.75)	46.0	0.142
4. I would like to learn this way in the future	6.58 (1.11)	6.25 (0.83)	49.5	0.204
5. I am interested in practicing more collaborative problem solving skills	6.42 (0.76)	5.83 (1.14)	50.0	0.215
6. I felt that the lesson was engaging	5.92 (1.17)	5.92 (1.04)	62.5	0.520
7. I found the lesson to be useful to me	5.92 (0.95)	5.92 (1.11)	70.0	0.928
8. I felt motivated to understand the activity	6.42 (0.95)	5.50 (0.87)	31.5	0.021*
9. I felt happy during the lesson	6.83 (0.37)	5.92 (1.19)	38.0	0.054
10. I felt excited during the lesson	6.83 (0.37)	5.67 (1.37)	31.0	0.019*
11. I felt bored during the lesson †	1.00 (0.0)	1.33 (0.62)	54.0	0.313
12. I felt confused during the lesson †	2.33 (1.5)	2.33 (1.7)	69.0	0.889
13. I felt sad during the lesson †	1.00 (0.0)	1.08 (0.28)	66.0	0.749
14. I felt scared during the lesson †	1.08 (0.28)	1.17 (0.37)	66.0	0.749
			z-score*	
Total	6.15 (1.56)	5.80 (1.49)	3.298	0.001**

*statistically significant $p= < 0.05$ **statistically significant $p= < 0.01$ † negatively coded question ***z-score calculation is referenced as the data was distributed normally. 1 = Strongly disagree / 7 = Strongly agree

Table 2. Results of the Presence Survey given to students after completing the CPS tasks.

Presence survey item	VR group *M (SD)*	DT group *M (SD)*	U-statistic	p-value
1. I forgot about my immediate surroundings while completing the problem solving activity	4.3 (1.9)	5.3 (0.75)	54	0.313
2. I was so involved in the problem solving activity I ignored everything else around me	4.8 (1.7)	5.8 (1.2)	49	0.194
3. I lost myself in this problem solving activity	3.3 (2.4)	3.9 (2.1)	60	0.509
4. I was so involved in the problem solving activity that I lost track of time	4.9 (1.8)	4.9 (2.0)	70.5	0.952
5. I blocked out things around me when I was completing the problem solving activity	4.2 (2.0)	5.6 (1.3)	42.5	0.095
6. When I was completing the problem solving activity, I lost track of this world around me	4.3 (1.8)	4.5 (1.0)	70	0.928
7. The time I spent completing the problem solving activity just slipped away	4.8 (2.1)	5.1 (1.6)	70	0.928
8. I was absorbed in the problem solving activity	5.7 (1.7)	5.5 (1.1)	59	0.484
9. During the problem solving activity I let myself go	5.3 (1.4)	4.9 (1.2)	58.5	0.453
10. I felt frustrated while completing this problem solving activity †	1.7 (1.5)	2.2 (1.1)	44	0.112
11. I found the problem solving activity confusing †	2.3 (1.4)	2.4 (1.8)	69.5	0.904
12. I felt annoyed while completing the problem solving activity †	1.0 (0.0)	1.4 (1.1)	48	0.174
13. I felt discouraged while completing the problem solving activity †	1.5 (0.89)	1.8 (1.2)	64.5	0.689
14. Completing the problem solving activity was tiring for my brain †	2.4 (1.66)	2.0 (1.4)	63	0.624
15. The problem solving activity was demanding	2.3 (1.42)	2.9 (1.0)	52.5	0.271
16. The problem solving activity was rewarding	5.3 (1.25)	5.1 (1.5)	67.5	0.818
17. I felt interested in the problem solving activity	6.9 (0.28)	6.0 (1.1)	34.5	0.032*
18. I felt involved in the problem solving activity	6.7 (0.62)	6.6 (0.6)	66.5	0.772
Total	5.48 (2.1)	5.50 (1.6)	1.168	0.242

*statistically significant $p = < 0.05$ † negatively coded question
1 = Strongly disagree / 7 = Strongly agree

Table 3. The number and type of observed CPS actions.

Collaborative Problem Solving Skills item	VR Group *M (SD)*	DT Group *M (SD)*	U-statistics	p-value
Collaborative Competence	243.83 (42.8)	207.83 (40.9)	10.5	0.263
(1) Establishing and maintaining shared understanding	137.17 (24.1)	115.50 (23.1)	7	0.093
(2) Taking appropriate action to solve the problem	96.33 (20.3)	82.50 (19.3)	11	0.298
(3) Establishing and maintaining team organisation	10.33 (7.6)	9.83 (6.3)	16	0.810
Problem Solving Process	247.33 (40.1)	195.00 (34.1)	6	0.066
(A) Exploring and understanding	141.50 (25.8)	106.33 (21.7)	6	0.066
(B) Representing and formulating	83.17 (18.9)	77.50 (10.7)	16.5	0.873
(C) Planning and executing	10.50 (8.7)	11.17 (8.9)	17.5	1.000
(D) Monitoring and reflecting	12.17 (4.8)	12.67 (11.5)	14.5	0.631
Total	491.17 (83.3)	314.17 (74.5)	0	0.005*

*statistically significant *p= < 0.05*
CPS (Collaborative Problem Solving) skills are derived from the Collaborative Competence and Problem Solving Process domains combined.

Table 4. Group time to complete CPS Activities in seconds

Time to complete activity in seconds' item	VR group *M (SD)*	DT group *M (SD)*	U-statistics	p-value
Consensus	384.17 (155.3)	330.33 (83.8)	12.00	0.379
Jigsaw	330.33 (103.1)	331.33 (89.6)	14.00	0.575
Total	730.17 (99.9)	671.67 (102.9)	14.00	0.575

Table 5. Rate of successful completion of task

Task success item	VR group *M (SD)*	DT group *M (SD)*	U-statistics	p-value
Jigsaw task 1	1 (0.0)	0.5 (0.5)	9	0.174
Jigsaw task 2	0.83 (0.38)	0.83 (0.3)	18	0.936
Total	1.83 (0.37)	1.33 (0.7)	11.5	0.337

Table 6. Number of words spoken to complete task

Number of words spoken item	VR group *M (SD)*	DT group *M (SD)*	U-statistics	p-value
Consensus	575.67 (209.8)	294.33 (140.4)	17.00	0.936
Jigsaw	558.67 (196.6)	516.67 (164.2)	15.00	0.689
Total	1134.33 (308.1)	811.00 (263.9)	8.00	0.129

All participants of the study responded to the Engagement and Presence Self-Report Survey. There were some problems with the recording of CPS instances due to the softly spoken nature of some of the students, and where this was the case, [inaudible] was noted to replace the word. This was not a problem for coding the CPS instance as the context of the sentences were not lost. None of the participants had previously participated in CPS activities of this nature.

The participants completed two questionnaires immediately after finishing the problem solving activity. These questionnaires asked participants to respond to a 7 point Likert-Type scale with 1 representing Strongly Disagree and 7 representing Strongly Agree. The data were analysed using a nonparametric two-tailed Mann-Whitney U test with a statistical significance level of 0.05. Table 1 shows the results of each participant group for the engagement survey, while Table 2 shows the results of the presence survey.

The engagement survey consisted of 14 questions (Table 1). Of these 14 questions, there are no statistically significant differences found between the VR group and the DT group for 12 responses. There is a significant difference found for the question "I felt motivated to understand the activity" with more positive views being reported from the VR group (M=6.42) compared with the Desktop group (5.50). There is also a statistically significant difference found between the VR group (M=6.83) and DT group (5.67) for the question "I felt excited during the lesson", again in favour of the VR group. The presence survey contained 18 questions (Table 2). Of these eighteen questions, there is no statistically significant difference between the two groups for all but one question. There is a statistically significant difference found for the question, "I felt interested in the problem solving activity" (VR group M=6.9, DT group M=6.0).

There is a statistically significant difference between the total scores of the engagement survey for the VR group (6.15) and DT group (5.80) in favour of the VR group. As there was a normal distribution of data, Table 1 records the z-score of a Mann-Whitney U Test. Unlike the engagement survey, there are no statistically significant differences between the VR group (5.48) and the DT group (5.48) when all answers to the presence survey were totalled. These data are not normally distributed, and therefore, Table 2 records the U Value.

Each groups CPS action was coded against guidance published in the CPS PISA Framework (OECD, 2015). Table 3 displays these results. An additional resource was created from the PISA Framework guidance to improve coding consistency and can be found in

Appendix 1. There is no statistically significant difference between the VR group and the DT group when comparing any single domain: Collaborative Competence or Problem Solving Process. However, there is a statistically significant difference between the VR group (491.17) and the DT group (M=314.17) when the Collaborative Competencies and Problem Solving Processes are combined to describe Collaborative Problem Solving skills (CPS skills). A two-tailed Mann-Whitney U test was conducted to reach these results.

In addition to the engagement and presence survey, and the recorded CPS skills three other data points were collected to examine differences between the VR group and DT group. These data along with the CPS skills data were coded and recorded after all the participants had completed their tasks using video and written transcript. The time for the participants to complete the single Consensus task and the combined time of the two Jigsaw tasks were recorded as well as the total time. There is no statistically significant difference between the VR group (M=384.17) and the DT group (M=330.33) for the Consensus task or the Jigsaw task (VR group M=330.33, DT group M=331.33). Table 3 displays these results. Again, no statistically significant time was evident for the combined times (VR group M=730.17, DT group =671.67). The successful completion of the Jigsaw tasks was recorded for each group and determined by an agreement between each pair upon the correct pin code as per an answer key. There is no statistically significant difference between the VR group (M=1.83) and the DT group (1.33) in these results. Table 4 displays these results. The number of words spoken was also recorded to determine differences between the VR group (M=1134.33) and DT group (M=811.00). Table 5 displays these results. There was also no statistical difference found in these data.

Discussion

The results of this study show that there are differences between collaborative problem solving activities completed in VR learning environments compared to activities completed using a desktop computer. The statistically significant differences appear in self-reported levels of engagement as well as in the total number of CPS skills used by participants to complete the activities. Students in the VR group reported a higher level of engagement in the activity. The VR group also demonstrated a larger number of CPS skills during the completion of activities. Although there were differences between the other measured areas, none of the differences were statistically significant.

The results of the engagement survey suggest that students engage more with CPS skill activities in VR learning environments. Although the results of only two individual questions show a statistically significant difference between the two groups, the combined total of the survey shows a statistical difference at $p=0.01$. It should be noted that the results show a positively engaged overall view towards the CPS activities for both groups. The results suggest that the VR learning environment, regardless of the interaction type, make these inherently engaging activities are even more engaging. This positive response from students in the VR group is in line with literature that suggests VR is a suitable learning environment for many different learning domains (Liu, Bhagat, Gao, Chang, & Huang, 2017).

VR group participants more strongly agreed at a statistically significant degree to the statement “I felt motivated to understand the activity” ($p=0.021$), compared with the DT group. There was also a statistically significant outcome in favour of the VR group for the statement, “I felt excited during the lesson” ($p=0.019$). This is additional evidence of the potential of VR learning environments as being valuable for more than just collaborative problem solving activities. As past studies have shown virtual reality field trips have been successfully used for educational purposes and the Consensus activity is arguably a comparable experience (Garcia-Cardona, Tian, & Prakoonwit, 2017). It is comparable because in the case of this experiment the students were situated in a room surrounded by hanging artwork, similar to that of an art gallery. When one compares this to the DT group activity, in which students interacted with the artwork on PowerPoint slides, the VR activity much more closely resembles a real-world art gallery. The increased level of excitement may be due to the immersive nature of this experience and the students having the feeling of ‘really being there’. Researchers have argued in the past that increased levels of immersion increases motivation, which may, lead to greater learning outcomes (Martirosov and Kopecek, 2017). This experiment also shows support for this argument. Researchers have also found that increased cognitive engagement in tasks leads to a willingness of students to put more effort into learning, which in part can also explain this result (Ding, Er, & Orey, 2018).

The study found that there was not a statistically significant difference between the two groups concerning the numbers of words spoken. However, what was clear was that the VR groups used more words to reach a Consensus on the order of paintings, while the DT Groups required more words to arrive at an agreed upon pin code in the Jigsaw tasks. Since the Jigsaw tasks had demonstrably correct answers, it is possible to argue that the optimal way to go about solving these problems was to arrive at the solution in the most efficient way possible, with the least number of words, and agreeing upon the correct solution. The VR group achieved better outcomes on both of these measures, with more correct responses overall and fewer words being used to reach these solutions. Although there was not a significant difference between the groups, a larger sample size could potentially yield more conclusive results. One reason for this difference may be due to the reduced number of environmental distractions encountered by the participants wearing the VR headset. The participants wearing the VR headset have the outside world and its inherent distractions occluded from view and replaced by the virtual world. Therefore, there are potentially less environmental distractions for those in the VR group compared with the DT group. Research from Bailenson et al. (2008) showed that task designers can control environmental distractions when building activities that utilise VR headsets to the benefit of learners. The potential for more significant exposure to environmental distractions for the DT group may explain why those students were less likely to come to an agreement on the correct pin code and use more words to do so. It should be noted that the VR learning environment is likely not in itself reducing the number of distractions, as it would just as easily be possible to build an intentionally distracting VR learning environment. In the case of this experiment, the virtual world was a plain classroom with very little rendered detail outside of four off-white classroom walls, a wooden coloured virtual table and the activity cards themselves. The DT group learning space was a typical classroom with windows to the outside world, visual displays on the walls and other typical classroom artefacts. It should be noted that the VR version of the activity had distractions, such as the unfamiliar controllers needed to interact with the objects and move within the 3D environment and perhaps the novel technology.

The number of words spoken during the Consensus task is higher for the VR group. Defining success for the Consensus task is more difficult due to its subjective nature. The success of the task only depends upon the participants reaching consensus on the final order of the paintings based on their preferences. Students on average will spend more time in seconds completing the Consensus task in the VR group in comparison with the DT group and use more words to reach this consensus. As a learning environment for developing CPS skills, this is arguably a desirable result as the students spend more time sharing their perspectives and taking into account the opinions of their partners, key aspects of CPS (Marzano & Heflebower, 2012).

The literature on immersion and presence state that higher levels of these phenomena can lead to higher levels of motivation for students to perform well (Webster, 2016). However, there was not a statistically significant difference on the immersion survey, and more careful research of this may be necessary to fully understand these results. Nevertheless, VR learning environments are arguably a more immersive medium and this may have lead to the students being more motivated to reach a genuine consensus, compared to those in the DT group who could be considered more likely to 'go through the motions'. Table 2 displays the one statistically significant answer on the Presence questionnaire "I felt interested in the problem solving activity", with the VR group (M=6.9) having more positive reactions to this statement than the DT group (M=6.0).

The potential advantages of using VR learning environments for engaging students in CPS activities is also evidenced by the higher number of recorded CPS skills observed during the experiment for the VR group (M=491.17) compared with the DT group (M=314.17). Furthermore, the mean number of Collaborative Competencies (VR group M=243.83, DT group M=207.83) and Problem Solving Processes (VR group M=247.33, DT group M=195.00) demonstrated were higher in VR; however, these were not at a level of statistical significance. Due to the higher number of Collaborative Competencies and Problem Solving Processes demonstrated in the VR learning environment, teachers are afforded more opportunities to provide feedback to their students in each domain. The sheer number of observable skills on their own may not be enough to argue that there is a benefit to using VR learning environments to teach CPS skills. However, combined with the evidence of the higher success rate of the Jigsaw tasks (VR group M=1.83, DT group M=1.33) as displayed in Table 5, the more in-depth discussion during the Consensus task and the statistically significant difference in the students overall level of engagement, VR learning environments seems to offer advantages over desktop tasks.

Some results of the study were surprising. As briefly mentioned already, the presence survey revealed no statistically significant difference between the VR group (M=5.48) and DT group (M=5.50). This result may be due to the fact that the VR task required students to interact with their partner who was not co-present in the virtual world. As the literature suggests, immersion in VR is linked to a subjects feeling that they are participating in a realistic experience (Romano & Brna, 2001). The fact participants of the VR group were unable to see their partner, may have deteriorated the level of immersion and erode the feeling of presence. Another way to explain the fact there was no statistically significant difference between the two groups could lie in the novelty of the technology. This experience was the first time any of the students had interacted in such a way within VR, and thus the new experience may have been overwhelming and made them all the more self-conscious of the experience.

Limitations of Research

The sample size used to conduct the study was limited to 12 students in each group. To draw more useful conclusions, the sample size needs to be larger and include a wider range of age groups and students from different backgrounds.

Although the research did not illuminate any profound differences between the VR group and the DT group, we can propose changes to the current study to potentially arrive at more statistically significant outcomes. Firstly, by researching either Consensus type tasks or Jigsaw type tasks, and not both task types simultaneously, it would potentially enable a closer look at what CPS skills present themselves. Secondly, the subjective nature of the CPS skills coding has serious drawbacks. The inter-rater reliability score is evidence of the limitation of subjective coding. Particularly with the adoption of PISA's CPS rubric, which was not intended to be used in such a way. With improvements in Artificial Intelligence, natural language processing, and learning analytics, it seems inevitable that such coding of behaviour will be able to be automated in the not too distant future. In the meantime, however, less advanced forms of quantitative data could be recorded to understand more reliably how each participant is acting. These data could be collected through automatically coding log-files of interactions with 3D assets within the learning environment. A similar strategy was employed to code CPS skills in the PISA 2015 CPS assessment (OECD, 2017).

Further research could improve the desktop task by moving the activity from the PowerPoint and into an interactive browser where the objects could be manipulated in an equal fashion to the VR environment. This change would have improved usability for the participants in the DT group, and provide an arguably better comparison between the two groups. Additionally, just as the Desktop environment did not take full advantage of the current technology available to activity designers, the VR environment itself does not take full advantage of what the medium has to offer. Research has shown that haptic feedback is a valuable addition to VR experiences to increase presence and immersion, which, as discussed in earlier sections of this paper, lead to more significant levels of engagement (Rnage, 2017). In defence of the current design of the research, the software used to create both the Desktop and VR activities required minimal technical skills and could be implemented by most teachers who use computers in their daily teaching practice if provided with the right hardware and basic guidance.

The measurement of student engagement is complex, and the use of questionnaires such as the one employed in this study have both advantages and disadvantages (Fredricks & McColskey, 2012). Biometric and neurometric measurements offer potentially objective measurements which could have been used to improve the validity of the results and provided more insight into the participants' experience with the CPS tasks. Examples of biometric data collection include Eye Tracking, which can measure visual attention, and Galvanic Skin Response which measures physiological arousal, while neurometric data can be collected using electroencephalography (EEG), which measures electrical activity within the brain (Schall, 2014). These technologies have matured to a point where they are relatively commonplace, especially within the realm of VR and educational technology research. There exist various headsets offering EEG and Eye Tracking data collection built into the headsets themselves with available data analysis (Soler-Dominguez, Camba, Contero, & Alcañiz, 2017, Tromp, Peeters, Meyer, & Hagoort, 2018).

Emerging technologies such as VR have the potential to disrupt teaching and learning. With the newest generation of high-end VR equipment finally reaching a price point that enables access to learning institutions outside of military organisations and universities, both research and development will accelerate rapidly. Once these VR technologies integrate with other emerging technologies, such as those mentioned in the previous two paragraphs, VR learning environments which would typically be too time-consuming, expensive or impossible to create, such as an environment to assess CPS skills, will become commonplace. The technology may not yet be there, and the price of such equipment is still too costly for the vast majority of institutions, however, just as the personal computer became a ubiquitous tool in learning spaces, so too will VR.

Acknowledgements

I would first like to thank my thesis advisor Senior Research Fellow of Educational Technology, Leo Aleksander Siiman (PhD), of the Faculty of Social Sciences, Institute of Education, at the University of Tartu. He provided the right amount of guidance, support and independence to ensure the writing and research process was a smooth one. I would also like to thank my thesis group members, Tija Briede and Niall Boyle who offered support, guidance and constructive criticism throughout the process as well as the other EduTech Masters students who participated in the online webinars and summer session in Tartu. Finally, I would like to thank Professor of Educational Technology, Margus Pedaste, of the University of Tartu and all of the other lectures and administrators of the EduTech Masters programme for the fruitful discussions we had on and off-line and the support they offered throughout the process.

Author's declaration

I hereby declare that I have written this thesis independently and that all contributions of other authors and supporters have been referenced. The thesis has been written in accordance with the requirements for graduation theses of the Institute of Education of the University of Tartu and is in compliance with good academic practices.

Signature: Date:

References

Assessment and Teaching of 21st Century Skills. Retrieved October 15, 2017, from http://www.atc21s.org/

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Michael, J. (2008). The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *The Journal of Learning Sciences*, *17*, 102–141. https://doi.org/0.1080/10508400701793141

Blanchard, C., Burgess, S., Halvill, Y., Lanier, J., Lasko, A., Oberman, M., & Teitel, M. (1990). Reality Built For Two: A Virtual Reality Tool. In *I3D '90* (pp. 35–36). New York, NY, USA: ACM. https://doi.org/10.1145/91385.91409

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the Linkages*. *Research in Higher Education*, *47*(1), 1–32. https://doi.org/10.1007/s11162-005-8150-9

Chang, C.-J., Chang, M.-H., Chiu, B.-C., Liu, C.-C., Fan Chiang, S.-H., Wen, C.-T., … Chen, W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education*, *114*, 222–235. https://doi.org/10.1016/j.compedu.2017.07.008

Cochrane, T., Cook, S., Aiello, S., Christie, D., Sinfield, D., Steagall, M., & Aguayo, C. (2017). A DBR Framework for Designing Mobile Virtual Reality Learning Environments. *Australasian Journal of Educational Technology*, *33*(Accepted for Special Issue on Mobile Augmented and Virtual Reality). https://doi.org/10.14742/ajet.3613

Collaborative Problem-Solving Framework PISA 2015. (2017, July). OECD.

de Oliveira Malaquias, F. F., & Malaquias, R. F. (2016). The role of virtual reality in the learning process of individuals with intellectual disabilities. *Technology & Disability*, *28*(4), 133–138. https://doi.org/10.3233/TAD-160454

Ding, L., Er, E., & Orey, M. (2018). An exploratory study of student engagement in gamified online discussions. *Computers & Education*, *120*, 213–226. https://doi.org/10.1016/j.compedu.2018.02.007

Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In

Handbook of Research on Student Engagement (pp. 763–782). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-2018-7_37

Garcia-Cardona, S., Tian, F., & Prakoonwit, S. (2017). Tenochtitlan - An Interactive Virtual Reality Environment that Encourages Museum Exhibit Engagement. In F. Tian, C. Gatzidis, A. El Rhalibi, W. Tang, & F. Charles (Eds.), *E-Learning and Games* (pp. 20–28). Springer International Publishing.

Gregory, S., Gregory, B., Reiners, T., Fardinpour, A., Hillier, M., Lee, M. J. W., … Larson, I. (2013). Virtual worlds in Australian and New Zealand higher education: Remembering the past, understanding the present and imagining the future. *Electric Dreams: 30th ASCILITE Conference*, 312–324.

Harman, J., Brown, R., & Johnson, D. (2017). Improved Memory Elicitation in Virtual Reality: New Experimental Results and Insights. In *Human-Computer Interaction - INTERACT 2017* (pp. 128–146). Springer, Cham. https://doi.org/10.1007/978-3-319-67684-5_9

Iskenderova, A., Weidner, F., & Broll, W. (2017). Drunk Virtual Reality Gaming: Exploring the Influence of Alcohol on Cybersickness. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 561–572). New York, NY, USA: ACM. https://doi.org/10.1145/3116595.3116618

Jensen, L., & Konradsen, F. (2017). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 1–15. https://doi.org/10.1007/s10639-017-9676-0

Kim, M., Lee, J., Jeon, C., & Kim, J. (2017). A Study on Interaction of Gaze Pointer-Based User Interface in Mobile Virtual Reality Environment. *Symmetry*, *9*(9), 189. https://doi.org/10.3390/sym9090189

Lele, A. (2013). Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

Liu, D., Bhagat, K., Gao, Y., Chang, T.-W., & Huang, R. (2017). *The Potentials and Trends of Virtual Reality in Education: A Bibliometric Analysis on Top Research Studies in the last Two Decades.*

Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers & Education*, *98*, 192–205. https://doi.org/10.1016/j.compedu.2016.03.018

Martirosov, S., & Kopecek, P. (2017). Virtual Reality and Its Influence on Training and Education - Literature Review. *Annals of DAAAM & Proceedings*, *28*, 708–717. https://doi.org/10.2507/28th.daaam.proceedings.100

Marzano, R. J., & Heflebower, T. (2012). *Teaching & Assessing 21st Century Skills*. Solution Tree Press.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Melo, M., Vasconcelos-Raposo, J., & Bessa, M. (2017). Presence and cybersickness in immersive content: Effects of content type, exposure time and gender. *Computers & Graphics*. https://doi.org/10.1016/j.cag.2017.11.007

Mitchell, I., & Carbone, A. (2011). A typology of task characteristics and their effects on student engagement. *International Journal of Educational Research*, *50*(5), 257–270. https://doi.org/10.1016/j.ijer.2011.05.001

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*(3), 598–610. https://doi.org/10.1037//0022-0663.94.3.598

O'Brien, H. L., & Toms, E. G. (2010). The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology*, *61*(1), 50–69.

OECD. (2017). *PISA 2015 Results (Volume V)*. OECD Publishing. https://doi.org/10.1787/9789264285521-en

P21 Framework Definitions. (2015, May). Partnership for 21st Century Learning. Retrieved from http://www.p21.org/storage/documents/docs/P21_Framework_Definitions_New_Logo_2015.pdf

Park, G. D., Allen, R. W., Fiorentino, D., Rosenthal, T. J., & Cook, M. L. (2006). Simulator Sickness Scores According to Symptom Susceptibility, Age, and Gender for an Older Driver Assessment Study. *Th ANNUAL MEETING*, 5.

Parmaxi, A., Stylianou, K., & Zaphiris, P. (2017). Leveraging Virtual Trips in Google Expeditions to Elevate Students' Social Exploration. In *Human-Computer Interaction*

– *INTERACT 2017* (pp. 368–371). Springer, Cham. https://doi.org/10.1007/978-3-319-68059-0_32

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000241

Rnage, F. (2017). Enhancing Virtual Reality experiences with feet haptics. Retrieved February 10, 2018, from http://publications.lib.chalmers.se/records/fulltext/251709/251709.pdf

Romano, D. M., & Brna, P. (2001). Presence and Reflection in Training: Support for Learning to Improve Quality Decision-Making Skills under Time Limitations. *CyberPsychology & Behavior*, *4*(2), 265–277. https://doi.org/10.1089/109493101300117947

Schall, A. (2014). New Methods for Measuring Emotional Engagement. In *Design, User Experience, and Usability. User Experience Design Practice* (pp. 347–357). Springer, Cham. https://doi.org/10.1007/978-3-319-07638-6_34

Siddiq, F., & Scherer, R. (2017). Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior*, *76*, 509–525. https://doi.org/10.1016/j.chb.2017.08.007

Soler-Dominguez, J. L., Camba, J. D., Contero, M., & Alcañiz, M. (2017). A Proposal for the Selection of Eye-Tracking Metrics for the Implementation of Adaptive Gameplay in Virtual Reality Based Games. In *Virtual, Augmented and Mixed Reality* (pp. 369–380). Springer, Cham. https://doi.org/10.1007/978-3-319-57987-0_30

Son, J., & Cho, Y. (2017). A Study on Capturing Learning Data from Virtual and Mixed Reality Cont.pdf. In *A Study on Capturing Learning Data from Virtual and Mixed Reality Contents Through Data Collection API*. Retrieved from http://www.icce2017.canterbury.ac.nz/sites/default/files/proceedings/workshops/W7/A%20Study%20on%20Capturing%20Learning%20Data%20from%20Virtual%20and%20Mixed%20Reality%20Cont.pdf

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, *50*(2), 862–869. https://doi.org/10.3758/s13428-017-0911-9

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review.

Computers in Human Behavior, *72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Vincur, J., Konopka, M., Tvarozek, J., Hoang, M., & Navrat, P. (2017). Cubely- Virtual Reality Block-Based Programming Environment.pdf. Presented at the VRST '17, Gothenburg, Sweden. https://doi.org/10.1145/3139131.3141785

von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a Collaborative Science Assessment Prototype. *Computers in Human Behavior*, *76*, 631–640. https://doi.org/10.1016/j.chb.2017.04.059

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, *24*(6), 1319–1333. https://doi.org/10.1080/10494820.2014.994533

Wu, T.-H., Wu, F., Liang, C.-J., Li, Y.-F., Tseng, C.-M., & Kang, S.-C. (2017). A virtual reality tool for training in global engineering collaboration. *Universal Access in the Information Society*, 1–13. https://doi.org/10.1007/s10209-017-0594-0

Zyngier, D. (2008). (Re)conceptualising student engagement: Doing education not doing time. *Teaching and Teacher Education*, *24*(7), 1765–1776. https://doi.org/10.1016/j.tate.2007.09.004

Appendices

Appendix 1. Adapted questionnaire: The development and evaluation of a survey to measure user engagement.

Likert-type scale: 1 Strongly Disagree - 7 Strongly Agree		
Original Question	New Question	Reason for change
I forgot about my immediate surroundings while shopping on this website.	I forgot about my immediate surrounds while in the lesson.	This better reflects the student experience.
I was so involved in my shopping task that I ignored everything around me.	I was so involved in my lesson that I ignored everything around me.	This better reflects the student experience.
I lost myself in this shopping experience.	I was very focused in the lesson.	Changed "lost myself to "was very focused" and "in this shopping experience" to "in the lesson". During the pre-study pilot, students were unsure of the expression "lost myself" so the change of phrase held the same meaning but was easier to understand. The second change better reflects the students' experience.
I was so involved in my shopping task that I lost track of time.	I was so involved in the lesson that I lost track of time.	This better reflects the student experience.
I blocked out things around me when I was shopping on this website.	I blocked out things around me when I was in the lesson.	This better reflects the student experience.
When I was shopping, I lost track of this world around me.	When I was in the lesson, I lost track of this world around me.	This better reflects the student experience.
The time I spent shopping just slipped away.	The time I spent in the lesson just slipped away.	This better reflects the student experience.
I was absorbed in my shopping task.	I was absorbed in my lesson.	No change
During this shopping experience I let myself go.	During the lesson I let myself go.	This better reflects the student experience.

Appendix 2.
Adapted questionnaire: Learning Science in Immersive Virtual Reality

Likert-type scale: 1 Strongly Disagree - 7 Strongly Agree			
	Original Questions	New Questions	Justification
1	"I used a lot of mental effort in the lesson"	No Change	I will conduct the whole experiment as if it is a stand alone lesson. So, there is no need to change
2	"I felt that the subject matter was difficult"	"I felt that the activity was difficult"	Change "subject matter" to "activity" to better reflect the student experience.
	"I have a good understanding of the material"	Remove from questionnaire	This question is not easily relatable to the activity.
3	"I enjoyed learning this way"	No Change	
4	"I would like to learn this way in the future"	No Change	
5	"I am interested in learning more about this subject"	"I am interested in practicing more collaborative problem solving skills"	Change "learning more about this subject" to "practicing more collaborative problem solving skills". This is to better reflect what we are interested in measuring.
6	"I felt that the lesson was engaging"	No change	
7	"I found the lesson to be useful to me"	No change	
8	"I felt motivated to understand the material"	"I felt motivated to understand the activity"	Change "material" to "activity" to better reflect the students experience.
9	"I felt happy during the lesson"	No change	
10	"I felt excited during the lesson"	No change	
11	"I felt bored during the lesson"	No change	
12	"I felt confused during the lesson"	No change	
13	"I felt sad during the lesson"	No change	

14	"I felt scared during the lesson"	No change	

Appendix 3. Guidance for the coding of collaborative problem solving skills and competencies.

Participant action	**Competence and Process**
Reading instructions card	A1, B2
Reading pin card numbers	A1, B2
Responding with yes/yeah to signify listening	A1
Asking a question about the order of the paintings	A1, A2
Answer questions about preferred painting / order of painting	B1
Statement about painting preference	B1, A1
Asking a clarifying question	A1, A2
Answering a clarifying question	A1, A2
Making a decision about the position or pin number position	B1, B2
Agree to position of painting or pin number position	B1, B2
Talking about mechanics of the problem solving process e.g. environment, how to place the cards etc.	A3
Descriptive statement about aesthetic of painting for judgement purposes e.g. *"Painting 1 has more skill and creativity than Painting 2."*	B1
Confirming or checking final order of painting or pin code	D2
Asking partner to wait	D2
Prompting partner to look around	A2
Responding with yes, no or other manner to signify listening to partner	A1
Checking if partner is listening	D3
Asking questions about the pin code order	A1, A2
Asking questions about what the partner can see	A1
Statement about pin order	A1, B2
Statement about roles of participants	A1, B3
Sharing information the participant can see	A1
Sharing information about individual action	C1
Asking partner to do something	C2

Asking what "we" should do	C3
Agreeing with plan	C3

Non-exclusive licence to reproduce thesis and make thesis public

I, Timothy Gallagher (date of birth: July 3, 1986),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright, 'Comparing collaborative problem solving in virtual reality and desktop computer display environments', supervised by Leo Aleksander Siiman (PhD).

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 06.06.2018